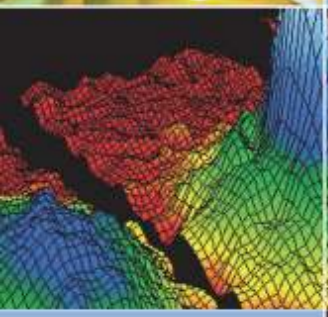
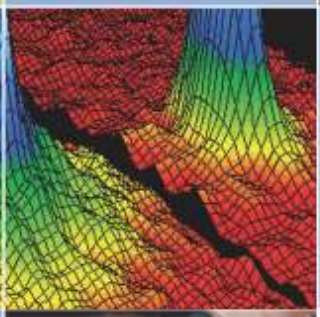



Collection
Les cours

Introduction au calcul des probabilités et à la statistique

Jean-François Delmas

ENSTA ParisTech

LES PRESSES DE L'ENSTA

Introduction au calcul des probabilités et à la statistique

Responsables de publication : Julien Bohdanowicz, Sophie Chouaf

En application du Code de la Propriété Intellectuelle et notamment de ses articles L. 122.4, L. 122-5 et L. 335-2, toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Une telle représentation ou reproduction constituerait un délit de contrefaçon, puni de trois ans d'emprisonnement et de 300 000 euros d'amende.

Ne sont autorisés que les copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations, sous réserve que soient indiqués clairement le nom de l'auteur et la source.

©*Les Presses de l'ENSTA, 2010*

Imprimé en France

ISBN 978-2-7225-0922-1

www.ensta.fr

Les cours

Jean-François Delmas

Introduction au calcul des probabilités et à la statistique

PARIS
LES PRESSES DE L'ENSTA
32, boulevard Victor, Paris 15^{e}
2010

À Ginger, Vickie et Gautier.

Préface

En écrivant ce livre, nous avons voulu présenter les outils élémentaires des probabilités et de la statistique mathématique avec, au travers des exercices, des applications sur des exemples concrets. Ce document est le support du cours "Introduction aux probabilités et à la statistique" de première année à l'École Nationale Supérieure des Techniques Avancées depuis 1999. Il est complété par un livre d'exercices et de problèmes corrigés.

Sans difficulté majeure, la lecture de ce livre nécessite la maîtrise des cours d'analyse des deux premières années du cycle de Licence ou des classes préparatoires.

Pour la théorie des probabilités, nous présentons les deux résultats fondamentaux suivants : la loi forte des grands nombres qui assure que la moyenne de données aléatoires converge vers la moyenne théorique quand le nombre d'observations indépendantes augmente et le théorème central limite (TCL) qui précise la vitesse de cette convergence. Nous donnons comme application importante l'estimation par intervalle de confiance. En statistique, nous présentons l'estimation paramétrique avec l'estimateur du maximum de vraisemblance, les régions de confiance et la théorie des tests avec en particulier le modèle de la régression linéaire et les tests d'adéquation du χ^2. En revanche, nous n'abordons pas la théorie de la statistique bayesienne.

En choisissant une présentation qui insiste sur les concepts des probabilités et de la statistique, nous avons décidé d'omettre certains résultats de la théorie de la mesure (ce qui par exemple nous oblige à traiter de manière apparemment différente le cas discret et le cas continu) et d'admettre les résultats techniques sur la transformation de Fourier. Nous admettons également les résultats sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance.

Résumons brièvement le contenu de ce livre. La première partie est consacrée à une introduction sur la théorie des probabilités. Le chapitre I donne la définition et les propriétés des probabilités. Il aborde en détail le cas des probabilités dis-

crètes avec quelques notions de dénombrement et des exemples de modélisation. Il présente également les notions de probabilités conditionnelles et d'indépendance. La notion de variable aléatoire ou résultat d'une expérience en présence d'aléa et la notion de moyenne ou espérance d'une variable aléatoire sont abordées au chapitre II. Ce chapitre est également consacré à l'étude des variables aléatoires discrètes, c'est-à-dire qui prennent un nombre fini ou dénombrable de valeurs. Dans le chapitre III, nous étudions les variables aléatoires à densité, qui prennent un continuum de valeurs, avec en particulier la loi gaussienne qui apparaît comme loi limite dans le TCL. La notion d'indépendance pour les variables aléatoires, qui joue un rôle crucial dans la loi forte des grands nombres et le TCL, est définie au paragraphe II.10 et détaillée pour les variables aléatoires discrètes (chapitre II) ou à densité (chapitre III). La notion d'espérance conditionnelle qui traduit la meilleure approximation, en un sens que nous préciserons, d'une variable aléatoire par une fonction d'une autre variable aléatoire donnée est définie pour les variables aléatoires discrètes (paragraphe II.14) ou à densité (paragraphe III.8). La problématique de la simulation est présentée au paragraphe III.9. La fonction caractéristique d'une loi, ou transformée de Fourier d'une mesure de probabilité, introduite dans le chapitre IV, est un outil important dans l'étude des variables aléatoires. Les théorèmes de convergence, avec les différents modes de convergence, sont abordés au chapitre V, avec en particulier la loi forte des grands nombres au paragraphe V.4 et le TCL au paragraphe V.6. Un exemple d'estimation par intervalle de confiance est détaillé au paragraphe V.7. Le chapitre VI est consacré aux vecteurs gaussiens qui apparaissent dans le TCL vectoriel comme lois limites et dans de nombreuses modélisations, comme par exemple la régression linéaire au paragraphe IX.7.

Enfin, la seconde partie est consacrée à la théorie de la statistique mathématique. Nous présentons au chapitre VII une introduction aux problématiques de la statistique mathématique. Le chapitre VIII traite de l'estimation de paramètres dans un modèle paramétrique. Ce chapitre aborde la construction d'estimateurs, dont l'estimateur du maximum de vraisemblance, et leurs propriétés avec les approches à horizon fini et asymptotique. Nous donnons les résultats sur la comparaison d'estimateurs pour le risque quadratique, la réduction de données à l'aide de statistiques exhaustives et l'amélioration d'estimateurs par conditionnement par rapport à une statistique exhaustive. Le chapitre IX est consacré à la théorie des tests. Nous introduisons en particulier au paragraphe IX.7 le modèle de la régression linéaire, c'est-à-dire l'explication d'une variable par une combinaison linéaire de variables explicatives et d'un bruit, et le test d'analyse de la variance pour l'utilité des variables explicatives. Nous présentons aussi plusieurs tests asymptotiques pour les modèles paramétriques. En particulier au paragraphe IX.9, nous introduisons les tests du χ^2 empiriques qui sont des tests d'adéquation de loi pour

des variables aléatoires prenant un nombre fini de valeurs, et nous terminons ce chapitre avec le paragraphe IX.10 sur quelques tests d'adéquation de loi non paramétriques. Dans le chapitre X, nous revenons sur les intervalles de confiance et leurs liens avec les tests. Enfin, le dernier chapitre XI fournit quelques tables de quantiles pour les lois usuelles.

Remerciements

Grâce à l'École Nationale Supérieure des Techniques Avancées et à ses étudiants, j'ai eu beaucoup de plaisir à enseigner ce cours. Je remercie Nicolas Bouleau pour m'avoir incité à faire cet enseignement et tous ceux qui y ont participé et ont contribué à son amélioration : Mohamed Ben Alaya, Hermine Biermé, Antoine Chambaz, Jean-Stéphane Dhersin, Anne Dutfoy, Xavier Epiard, Marie-Pierre Etienne, Josselin Garnier, Julien Guyon, Far Hadda, Lauris Joubert, Régis Lebrun, Vincent Lefieux, Jérôme Lelong, Eulalia Nualart, Christian Paroissin, Bénédicte Puig, Victor Rivero, Raphaël Roux, Mohamed Sbaï, Simone Scotti, Michel Sortais, Emmanuel Temam et Mathias Winkel.

Vraisemblablement, je n'aurais pas pris autant de plaisir à rédiger ce livre sans les nombreuses discussions et digressions avec Benjamin Jourdain et Bernard Lapeyre en probabilité et avec Didier Chauveau, Eric Parent et Jean-Pierre Raoult en statistique. Je remercie aussi Jean-Philippe Chancelier pour son aide précieuse concernant l'utilisation des logiciels Latex pour le texte et surtout Scilab (`http://www.scilab.org/`) pour les illustrations. Enfin je remercie tous mes collègues du Cermics, laboratoire de l'École des Ponts et Chaussées, pour l'ambiance de travail agréable et stimulante qu'ils ont su créer et développer.

Gardons pour la fin ceux qui m'entourent au plus près et m'ont donné dès le début d'autres définitions de l'espérance.

Champs sur Marne,
Octobre 2009. *Jean-François Delmas*

Un bref historique

Le calcul des probabilités

"Un problème relatif aux jeux de hasard proposé à un austère janséniste par un homme du monde a été à l'origine du calcul des probabilités" Denis Poisson (1781-1840). Le Chevalier de Méré proposa à Blaise Pascal (1623-1662) des problèmes sur les jeux de hasard dont le "problème des parties" : Le prix d'un tournoi est gagné par le premier participant qui remporte un nombre fixé de parties. Si l'on interrompt le jeu avant la fin, comment répartir équitablement le prix entre les participants ? De nombreuses solutions fausses avaient été proposées pour ce problème vieux de deux siècles. Pascal en donna une solution correcte qu'il soumit à Pierre de Fermat (1601-1665) en 1654. Il publia sa solution dans son "Traité du triangle arithmétique" en 1665.

En 1657, le livre "De ratiociniis in ludo aleae" de Christiaan Huygens (1629-1695) exposa les concepts fondamentaux du calcul des probabilités comme le calcul de l'espérance d'une variable aléatoire prenant un nombre fini de valeurs.

Dans son ouvrage posthume "Ars conjectandi" en 1713, Jacques Bernoulli (1654-1705) approfondit les résultats de Huygens. Il démontra aussi, à l'aide du calcul combinatoire, la loi des grands nombres (convergence de la moyenne empirique vers la moyenne) qui fut à l'origine de l'essor des probabilités. En 1733, dans "The doctrine of chances", Abraham de Moivre (1667-1754) précisa dans un cas particulier la vitesse de convergence de la loi des grands nombres ; ce fut la première version du théorème central limite. Ce résultat fut étendu par Pierre-Simon Laplace (1749-1827). Ce dernier en utilisant le calcul infinitésimal et en développant les fonctions génératrices et les fonctions caractéristiques dans "Théorie analytique des probabilités", paru en 1812, dépassa le cadre du calcul combinatoire et donna un nouvel élan au calcul des probabilités.

Les résultats généraux sur la loi des grands nombres et le théorème central limite furent établis au XIX^{e} siècle par Denis Poisson (1781-1840), Irénée-Jules Bienaymé

(1796-1878) et l'école de St Petersbourg avec Pafnouti Tchebychev (1821-1894), Andreï Markov (1856-1922) et Alexandre Liapounov (1857-1918).

Au XX^{e} siècle, la théorie de la mesure et de l'intégration permit de clarifier les notions du calcul des probabilités : mesures de probabilité, variables aléatoires, lois, espérances, lois conditionnelles. La monographie d'Andreï Kolmogorov (1903-1987) "Grundbegriffe der Wahrscheinlichkeitsrechnung" parue en 1933 donna le cadre théorique dans lequel s'exprime encore aujourd'hui le calcul des probabilités.

Dès la première moitié du XX^{e} siècle, le calcul des probabilités connaît un nouvel essor avec l'étude des processus stochastiques et surtout leurs nombreuses applications. Celles-ci se sont multipliées dans la deuxième moitié du siècle : modélisation de phénomènes physiques (en particulier au niveau microscopique pour les fluides complexes ou les matériaux et en physique statistique) ou biologique (en démographie et épidémiologie, mais aussi au niveau de la cellule ou de l'ADN), en informatique (analyse d'algorithmes, d'images ou de réseaux), en économie (assurance ou finance de marché) ainsi que dans les sciences de l'ingénieur (fiabilité, optimisation, analyse de risque, maîtrise de l'environnement aléatoire). Enfin, avec la puissance accrue des ordinateurs, les simulations et les méthodes de Monte-Carlo, développées dans les années 1940, ont amplifié l'utilisation des modèles aléatoires et sont devenues un domaine important des probabilités.

La statistique

Le mot "statistique" vient de l'allemand "Statistik", qui, au milieu du XVII^{e} siècle, désigne l'analyse des données utiles à l'État. Le traitement d'un grand nombre de données chiffrées qui sont triées, classées ou résumées correspond à ce que l'on appelle aujourd'hui "les statistiques" au pluriel. On les distingue de "la statistique", au singulier, qui correspond à la modélisation de ces données, vues comme résultats d'expériences en présence d'aléa, et à l'étude de cet aléa.

On peut dater l'émergence de la statistique du début du XIX^{e} siècle, avec l'étude de données provenant de l'astronomie sur les positions des planètes et leur trajectoire. En particulier, en 1805 Adrien-Marie Legendre (1752-1832) introduisit la méthode des moindres carrés pour estimer des coefficients à partir de données, et en 1809 Carl Friedrich Gauss (1777-1855), utilisant une modélisation des erreurs par la loi normale, retrouva en maximisant la densité de la loi normale des erreurs (i.e. la vraisemblance ou loi *a posteriori*) l'estimation par moindres carrés. Ces travaux influencèrent Pierre-Simon Laplace (1749-1827) qui en 1810 montra que la loi normale apparaît naturellement comme loi des erreurs grâce au théorème central limite. Dans son livre sur "l'homme moyen" en 1835, Adolphe Quetelet (1796-1874) utilisa les résultats de Laplace pour analyser des données sociales à l'aide de la loi normale, et montrer la stabilité de ces données sur plusieurs années.

Il faut attendre la fin du XIX^{e} siècle pour une nouvelle avancée dans le domaine de la statistique. En 1885, Francis Galton (1822-1911) présenta une étude sur la taille des garçons en fonction de la taille moyenne des parents. Il observa à la fois un phénomène de dépendance, qui sera traduit par un effet de corrélation, et de retour à la moyenne ou régression. Karl Pearson (1857-1936) et Udny Yule (1871-1951), à partir des travaux de Francis Edgeworth (1845-1926) sur les lois normales multidimensionnelles, étendirent la régression linéaire à un cadre plus général. Il faut également souligner les tests d'adéquation du χ^2 introduits par Pearson en biométrie à la fin du XIX^{e} siècle.

Au début du XX^{e} siècle, les années 1920 sont marquées par les travaux fondamentaux de Ronald Fisher (1890-1962) qui sont motivés par des problèmes d'agronomie. Fisher introduisit en particulier les notions de modèle statistique, d'exhaustivité et d'estimateur du maximum de vraisemblance. L'utilisation de modèles statistiques permit ainsi d'analyser des données peu nombreuses. Signalons également sur le même thème les résultats de William Gosset (1876-1937) pour les échantillons gaussiens. Travaillant pour la brasserie Guiness, il prit le pseudonyme de Student pour publier ses travaux.

Motivé par l'étude des effets de différents traitements en agriculture, Jerzy Neyman (1894-1981) introduisit en 1934 puis développa avec Egon Pearson (1895-1980) l'estimation par intervalles de confiance et les tests d'hypothèses. Le nom d'hypothèse "nulle" provient de l'hypothèse correspondant à l'absence d'effet du traitement considéré. En 1940, Abraham Wald (1902-1950) proposa une vision unifiée de la théorie de l'estimation et des tests d'hypothèses.

À partir des années 1950, la statistique connaît une croissance exponentielle avec des applications dans tous les domaines : sciences de l'ingénieur, sciences expérimentales, sciences sociales, médecine et sciences du vivant, économie, ... Elle est devenue un outil incontournable pour l'analyse et la compréhension des données.

Table des matières

partie I Calcul des probabilités

I Espaces probabilisés **3**
- I.1 Vocabulaire . . . 3
- I.2 Probabilités . . . 4
- I.3 Probabilités sur un ensemble fini ou dénombrable . . . 7
- I.4 La modélisation (I) . . . 9
- I.5 Dénombrement . . . 10
- I.6 Probabilités conditionnelles . . . 12
- I.7 Indépendance . . . 13
- I.8 Modélisation (II) . . . 14
- I.9 Rappels sur les ensembles . . . 15
- I.10 Compléments sur les espaces mesurables et les fonctions mesurables 15
- I.11 Résumé . . . 18
- I.12 Exercices . . . 19

II Variables aléatoires discrètes **25**
- II.1 Variables aléatoires . . . 26
- II.2 Exemples de variables aléatoires discrètes . . . 27
- II.3 Loi d'un vecteur, lois marginales . . . 29
- II.4 Variables aléatoires discrètes indépendantes (I) . . . 31
- II.5 Schéma de Bernoulli et autres exemples . . . 32
- II.6 Changement de variable . . . 39
- II.7 Espérance d'une variable aléatoire quelconque . . . 39
- II.8 Espérance d'une variable aléatoire discrète . . . 44
- II.9 Variance et Covariance . . . 47
- II.10 Indépendance (II) . . . 48
- II.11 Loi faible des grands nombres . . . 51

II.12 Fonctions génératrices 52
II.13 Indépendance (III) 55
II.14 Lois conditionnelles et espérances conditionnelles 56
II.15 Rappels sur les séries et les séries entières 62
II.16 Résumé 64
II.17 Exercices 67

III Variables aléatoires à densité 79
III.1 Définitions 80
III.2 Lois marginales 83
III.3 Espérance 84
III.4 Lois usuelles 85
III.5 Autres lois 88
III.6 Indépendance 90
III.7 Calcul de lois 93
III.8 Lois conditionnelles 95
III.9 Simulation 97
III.10 Rappels sur l'intégration 99
III.11 Résumé 103
III.12 Exercices 106

IV Fonctions caractéristiques 113
IV.1 Définitions 113
IV.2 Propriétés 115
IV.3 Fonctions caractéristiques usuelles 118
IV.4 Résumé 121
IV.5 Exercices 123

V Convergences et théorèmes limites 125
V.1 Convergence presque sûre et théorèmes limites 125
V.2 Convergence en probabilité et dans l'espace L^2 128
V.3 Convergence en loi 132
V.4 Loi forte des grands nombres 137
V.5 Estimations de lois 141
V.5.1 Variables aléatoires discrètes 141
V.5.2 Variables aléatoires réelles 141
V.5.3 Variables aléatoires à densité 142
V.6 Théorème central limite 145
V.7 Autour du théorème central limite (I) 149
V.8 Autour du théorème central limite (II) 153
V.9 Résumé 155

V.10 Exercices 158

VI Vecteurs gaussiens **165**
VI.1 Définition et propriétés 165
VI.2 Loi du χ^2, loi de Student, loi de Fisher 173
VI.3 Théorème central limite vectoriel 179
VI.4 Résumé 181
VI.5 Exercices 183

partie II Statistique

VII Introduction à la statistique : un exemple **189**
VII.1 Estimation ponctuelle 189
VII.2 Test d'hypothèses 190
VII.3 Intervalle de confiance 191

VIII Estimation ponctuelle **195**
VIII.1 Hypothèses sur le modèle 195
VIII.2 Statistiques et estimateurs 197
VIII.3 Construction d'estimateurs convergents 198
VIII.3.1 Méthode de substitution 198
VIII.3.2 Méthode des moments 198
VIII.3.3 Le maximum de vraisemblance 199
VIII.4 Choix d'un estimateur 204
VIII.4.1 Risque quadratique et comparaison d'estimateurs 204
VIII.4.2 Score, information de Fisher, modèle régulier 207
VIII.4.3 Borne FDCR 210
VIII.4.4 Modèle gaussien 212
VIII.5 Amélioration d'estimateurs 214
VIII.5.1 Statistiques exhaustives, statistiques totales 214
VIII.5.2 Estimateurs améliorés de Rao-Blackwell 216
VIII.5.3 Le modèle exponentiel 218
VIII.6 Analyse asymptotique 222
VIII.6.1 Estimateurs de substitution 223
VIII.6.2 Estimateurs des moments 223
VIII.6.3 Estimateurs du maximum de vraisemblance 224
VIII.6.4 Comparaison asymptotique 226
VIII.7 Résumé 228
VIII.8 Exercices 231

IX Tests d'hypothèses **235**
IX.1 Tests . . . 236
IX.2 Erreurs . . . 237
IX.3 Choix d'un test . . . 238
IX.4 Test d'hypothèses simples . . . 239
IX.5 Statistique de test et p-valeur . . . 242
IX.6 Hypothèses composites pour les modèles exponentiels . . . 244
IX.7 Régression linéaire . . . 248
IX.7.1 Modèle et estimation . . . 248
IX.7.2 Test d'utilité des régresseurs . . . 251
IX.8 Tests asymptotiques . . . 255
IX.8.1 Définitions et exemples . . . 255
IX.8.2 Hypothèse implicite : le test de Wald . . . 260
IX.8.3 Hypothèse explicite : le test de Hausman . . . 264
IX.9 Test d'adéquation du χ^2 et applications . . . 267
IX.9.1 Test du χ^2 empirique . . . 267
IX.9.2 Test d'adéquation à une loi . . . 270
IX.9.3 Test d'indépendance . . . 271
IX.9.4 Test du χ^2 empirique (démonstration) . . . 273
IX.10 Autres tests asymptotiques . . . 275
IX.10.1 Test de Kolmogorov-Smirnov pour un échantillon . . . 275
IX.10.2 Test de Kolmogorov-Smirnov pour deux échantillons . . . 278
IX.10.3 Test de comparaison pour deux échantillons . . . 279
IX.11 Résumé . . . 282
IX.12 Exercices . . . 289

X Régions de confiance, Intervalles de confiance **293**
X.1 Régions et intervalles de confiance de niveau exact . . . 293
X.2 Régions et intervalles de confiance de niveau approché . . . 296
X.2.1 Niveau par excès . . . 296
X.2.2 Niveau asymptotique . . . 298
X.3 Régions de confiance et tests . . . 301
X.4 Résumé . . . 302
X.5 Exercices . . . 303

XI Tables statistiques **305**
XI.1 Quantiles de la loi $\mathcal{N}(0,1)$ . . . 305
XI.2 Fonction de répartition de la loi $\mathcal{N}(0,1)$ . . . 306
XI.3 Quantiles de la loi du χ^2 . . . 307
XI.4 Quantiles de la loi de Student . . . 308
XI.5 Quantiles de la loi de Fisher-Snedecor . . . 309

Références **311**

Index **313**

Première partie

Calcul des probabilités

I

Espaces probabilisés

Ce chapitre est consacré à la définition et aux propriétés des probabilités (pour une introduction générale à la théorie de la mesure, le lecteur pourra consulter les ouvrages [12] ou [14]).

Après la brève introduction au vocabulaire usuel en probabilité du paragraphe I.1, on donne la définition et quelques propriétés des probabilités au paragraphe I.2. On étudie plus en détail le cas des probabilités sur un espace fini ou dénombrable aux paragraphes I.3 et I.5. La notion de probabilité conditionnelle est introduite au paragraphe I.6, et la notion importante de l'indépendance au paragraphe I.7. Les paragraphes I.9 et I.10 donnent quelques rappels sur les opérations d'ensemble et des compléments sur la théorie de la mesure. Enfin, les paragraphes I.4 et I.8 sont consacrés à la modélisation du lancer de deux dés à l'aide du formalisme des probabilités.

I.1 Vocabulaire

Afin d'illustrer le vocabulaire et les notations probabilistes, on considère l'exemple du lancer d'un dé à 6 faces.

Vocabulaire usuel	vocabulaire consacré	notation	exemple
un résultat possible	une réalisation	ω	5
tous les résultats possibles	l'espace des réalisations ou espace d'états	Ω	$\{1,2,3,4,5,6\}$
un sous-ensemble de résultats possibles	un évènement	$A \subset \Omega$	$\{2,4,6\}$

On définit également le vocabulaire suivant.

l'évènement certain	Ω
l'évènement impossible	$\emptyset$
l'évènement contraire de A	A^c
l'évènement A et B	$A \cap B$
l'évènement A ou B (non exclusif)	$A \cup B$

On dit que les évènements A et B sont incompatibles si $A \cap B = \emptyset$.

I.2 Probabilités

Intuitivement, pour estimer la probabilité d'obtenir un 5 ou un 6 lors d'un lancer d'un dé à 6 faces, i.e. la probabilité de l'évènement $A = \{5, 6\}$ que l'on notera $\mathbb{P}(A)$, on effectue un grand nombre, N, de lancers du même dé, et on compte le nombre $N(A)$ de fois où le résultat est 5 ou 6. La fréquence empirique des succès, $\dfrac{N(A)}{N} \approx \dfrac{1}{3}$ (si le dé n'est pas biaisé), est une bonne approximation de $\mathbb{P}(A)$. La figure I.1 représente le résultat d'une simulation de 1000 lancers, où l'on a représenté chaque apparition de 5 ou 6 par un trait vertical. Enfin on observe que la fréquence empirique $N(A)/N$ "converge" vers $1/3$ quand N augmente.

On verra au chapitre V sur les théorèmes limites comment justifier cette approximation à l'aide de la loi forte des grands nombres. De cette approche intuitive, on peut induire des relations satisfaites par la probabilité $\mathbb{P}$. Comme $N(\Omega) = N$ et $N(\emptyset) = 0$, on doit donc avoir $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset) = 0$. Si on considère des évènements disjoints (incompatibles) A et B, on a $N(A \cup B) = N(A) + N(B)$, il est donc naturel d'imposer $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ si $A \cap B = \emptyset$.

Si Ω est au plus dénombrable, on note dans ce cas $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble des sous-ensembles de Ω. De manière plus générale on considère un ensemble $\mathcal{F} \subset \mathcal{P}(\Omega)$ contenant Ω, stable par passage au complémentaire et stable par réunion dénombrable (cf. définition I.12). On dit que $\mathcal{F}$ est une tribu et $(\Omega, \mathcal{F})$ un espace mesurable (voir le paragraphe I.10 pour des exemples). Dans une première lecture, on pourra supposer que les espaces que l'on manipule sont finis ou dénombrables. En fait les paragraphes I.2, I.6 et I.7 couvrent le cas général. On dit que $A \subset \Omega$ est un **évènement** si $A \in \mathcal{F}$.

Définition I.1. *Une **probabilité** (ou une mesure de probabilité) est une fonction $\mathbb{P}$ de l'ensemble des évènements $\mathcal{F}$ vers $[0, 1]$ telle que :*

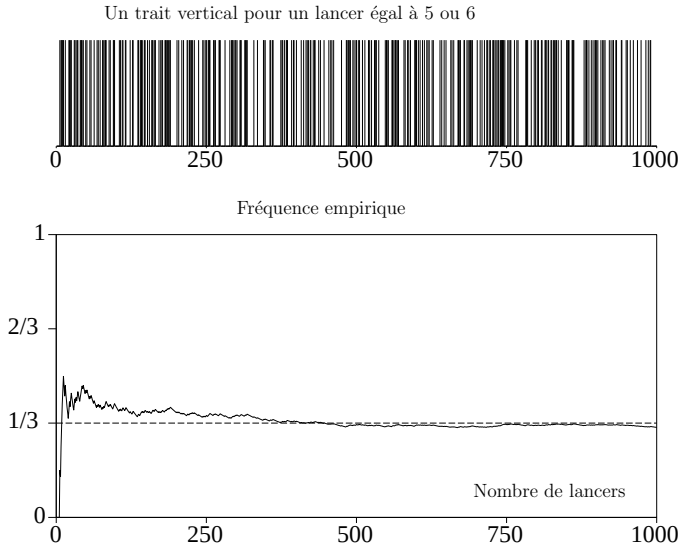


Figure I.1. Apparitions de 5 ou 6 lors de 1000 lancers d'un dé.

- $\mathbb{P}(\Omega) = 1$ *et* $\mathbb{P}(\emptyset) = 0$.
- *Si* $(A_i, i \in I)$ *est une collection* ***finie*** *ou* ***dénombrable*** *d'évènements* ($A_i \in \mathcal{F}$ *pour tout* $i \in I$) ***disjoints*** *deux à deux, alors :*

$$\boxed{\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbb{P}(A_i).} \tag{I.1}$$

Cette propriété s'appelle la σ***-additivité****. (On parle d'additivité si l'ensemble d'indices* I *est fini.)*

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé **espace probabilisé** ou espace de probabilités. On dit que l'évènement A est **presque sûr** (noté p.s.) si $\mathbb{P}(A) = 1$. On dit également que l'évènement A est **négligeable** si $\mathbb{P}(A) = 0$.

Exemple. Pour la modélisation du lancer de dé équilibré, on choisit $\Omega = \{1, \ldots, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$. ♢

Exemple. On considère $\Omega = [0, 1]$, et on choisit pour $\mathcal{F}$ la tribu borélienne sur $[0, 1]$ notée $\mathcal{B}([0, 1])$ (voir le paragraphe I.10 et la remarque I.13). On peut définir la probabilité d'un borélien A par sa mesure de Lebesgue $\mathbb{P}(A) = \lambda(A)$. En particulier si $0 \leq a \leq b \leq 1$, on a $\mathbb{P}([a, b]) = b - a$. ♢

On renvoie au paragraphe II.15 pour les propriétés des sommes dénombrables de termes positifs qui apparaissent par exemple dans l'égalité (I.1). On donne quelques propriétés des probabilités.

Proposition I.2. *Soit A et B deux évènements.*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
3. *Si $A \subset B$ alors on a $\mathbb{P}(A) \leq \mathbb{P}(B)$ (propriété de monotonie).*
4. *Soit $(A_i, i \in I)$ une collection finie ou dénombrable d'évènements disjoints deux à deux tels que $\sum_{i\in I} \mathbb{P}(A_i) = 1$, on a :*

$$\boxed{\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(A_i \cap B)} \quad \textit{(formule de décomposition).} \tag{I.2}$$

5. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'évènements ($\forall m \geq n \geq 0, A_n \subset A_m$). Alors on a :*

$$\boxed{\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)} \quad \textit{(convergence monotone).}$$

La formule du crible (I.6) (cf. exercice I.8) est une généralisation de la propriété 2 pour n évènements.

Démonstration. On remarque que $A^c \in \mathcal{F}$ par la propriété 2 de la définition I.12. On a $\Omega = A \cup A^c$ et les évènements A et A^c sont disjoints. Donc par additivité de $\mathbb{P}$, on a :

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1.$$

Cela implique la première égalité. Pour la deuxième, on décompose $A \cup B$ comme la réunion de trois ensembles disjoints : $A \cap B^c$, $A \cap B$ et $A^c \cap B$. Par additivité, on obtient alors :

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)\\ &= [\mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B)] + [\mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B)] - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).\end{aligned}$$

Pour l'inégalité du 3, on pose $C = A^c \cap B \in \mathcal{F}$. A et C sont disjoints donc par additivité, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(C)$. Comme $\mathbb{P}(C) \in [0, 1]$, on en déduit que $\mathbb{P}(C)$ est positif, et donc $\mathbb{P}(B) \geq \mathbb{P}(A)$.

Pour démontrer la propriété 4, on note $A = \left(\bigcup_{i\in I} A_i\right)^c$ de sorte que $\Omega = A \cup \left(\bigcup_{i\in I} A_i\right)$. De plus l'évènement A est négligeable. En effet, par σ-additivité, on a $\mathbb{P}(A) = 1 - \sum_{i\in I} \mathbb{P}(A_i) = 0$. En utilisant (I.5) puis la σ-additivité, il vient :

$$\begin{aligned}\mathbb{P}(B) = \mathbb{P}(B\cap\Omega) &= \mathbb{P}\left(B \cap \left(A \cup \bigcup_{i\in I} A_i\right)\right)\\ &= \mathbb{P}\left((B\cap A) \cup \bigcup_{i\in I}(B\cap A_i)\right)\\ &= \mathbb{P}(B\cap A) + \sum_{i\in I}\mathbb{P}(B\cap A_i).\end{aligned}$$

Par monotonie, on a $\mathbb{P}(B\cap A) \leq \mathbb{P}(A) = 0$. On en déduit donc que $\mathbb{P}(B) = \sum_{i\in I}\mathbb{P}(B\cap A_i)$.

On démontre la propriété 5. On pose $B_0 = A_0$ et pour tout entier $n \geq 1$, $B_n = A_n \cap A_{n-1}^c \in \mathcal{F}$. On a $\bigcup_{n\leq N} B_n = A_N$ et $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$. De plus les évènements $(B_n, n \geq 0)$ sont disjoints deux à deux. Donc par σ-additivité, on a $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}}\mathbb{P}(B_n)$. Pour conclure il suffit de remarquer que la somme partielle $\sum_{n\leq N}\mathbb{P}(B_n) = \mathbb{P}(\bigcup_{n\leq N} B_n) = \mathbb{P}(A_N)$ converge en croissant vers $\sum_{n\in\mathbb{N}}\mathbb{P}(B_n) = \mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n)$, car les termes sont positifs.

□

I.3 Probabilités sur un ensemble fini ou dénombrable

On se concentre maintenant sur le cas où Ω est au plus dénombrable. On rappelle qu'alors, on considère $\mathcal{F} = \mathcal{P}(\Omega)$. Dans ce cas la probabilité $\mathbb{P}$ est entièrement déterminée par la collection $(\mathbb{P}(\{\omega\}), \omega \in \Omega)$ comme l'indique le lemme suivant.

Lemme I.3. *Pour tout $A \in \mathcal{F}$, on a* $\boxed{\mathbb{P}(A) = \sum_{\omega\in A}\mathbb{P}(\{\omega\}).}$

Démonstration. Les évènements $(A\cap\{\omega\}, \omega\in\Omega)$ sont disjoints deux à deux et par σ-additivité, on a :

$$\mathbb{P}(A) = \mathbb{P}(A\cap\Omega) = \mathbb{P}\left(\bigcup_{\omega\in\Omega}(A\cap\{\omega\})\right) = \sum_{\omega\in\Omega}\mathbb{P}(A\cap\{\omega\}).$$

On aurait pu bien sûr appliquer la formule de décomposition de la proposition I.2. On remarque ensuite que si $\omega \in A$ alors $A \cap \{\omega\} = \{\omega\}$ et $\mathbb{P}(A \cap \{\omega\}) = \mathbb{P}(\{\omega\})$, et si $\omega \notin A$, alors $\mathbb{P}(A \cap \{\omega\}) = \mathbb{P}(\emptyset) = 0$. On en déduit que :

$$\sum_{\omega \in \Omega} \mathbb{P}(A \cap \{\omega\}) = \sum_{\omega \in A} \mathbb{P}(A \cap \{\omega\}) + \sum_{\omega \in A^c} \mathbb{P}(A \cap \{\omega\}) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

□

Un exemple important de probabilité sur un ensemble fini est la probabilité uniforme.

Définition I.4. *Si Ω est fini et si pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = 1/\text{Card } \Omega$, alors on dit que $\mathbb{P}$ est la probabilité* ***uniforme*** *sur Ω.*

Il est immédiat de déduire du lemme I.3 le résultat suivant.

Corollaire I.5. *Soit Ω fini et $\mathbb{P}$ la probabilité uniforme sur Ω. Pour tout $A \subset \Omega$, on a :*

$$\boxed{\mathbb{P}(A) = \frac{\text{Card } A}{\text{Card } \Omega}.} \tag{I.3}$$

Sous la probabilité uniforme, la probabilité d'un évènement correspond au nombre de cas favorables (i.e. le cardinal de l'évènement) divisé par le nombre de cas possibles (i.e. le cardinal de Ω).

Exercice I.1.
On tire au hasard deux cartes dans un jeu de 52 cartes.

1. Quelle est la probabilité pour que la couleur des deux cartes soit pique ?
2. Quelle est la probabilité pour que les deux cartes ne soient pas de la même couleur (pique, cœur, carreau, trèfle) ?
3. Quelle est la probabilité pour que la première carte soit un pique et la seconde un cœur ?
4. Quelle est la probabilité pour qu'il y ait un pique et un cœur ?
5. Quelle est la probabilité pour qu'il y ait un pique et un as ?

△

▼ *Correction* I.1. 1) $\frac{1}{17}$; 2) $\frac{13}{17}$; 3) $\frac{13}{52}\frac{13}{51} = \frac{13}{204}$; 4) $\frac{13}{102}$; 5) $\frac{2}{52} + 2\frac{12}{52}\frac{3}{51} = \frac{29}{26 * 17}$.

I.4 La modélisation (I)

Il est souvent très délicat de choisir le modèle qui représente la réalité. Par exemple en l'absence d'information on choisit généralement la probabilité uniforme sur l'ensemble des réalisations. Le problème peut toutefois devenir rapidement compliqué. On traite maintenant un cas simple, mais on pourra aussi consulter les exercices I.15 et I.21. On veut donner un modèle probabiliste pour la somme des faces d'un lancer de deux dés équilibrés. L'espace des résultats est donc $\Omega_1 = \{2, 3, \ldots, 12\}$. Quelle probabilité sur Ω_1 modélisera au mieux cette expérience ? Est-ce la probabilité uniforme $\mathbb{P}_1$? Dans ce cas, on a $\mathbb{P}_1(\{7\}) = \mathbb{P}_1(\{12\}) = 1/11$, mais cela n'est pas conforme à l'expérience. En effet 12 est beaucoup plus rare que 7. De plus on ne tient pas compte de l'information suivante : le résultat est la somme de deux dés. L'idée est donc de considérer l'espace d'états correspondant aux résultats de chacun des dés : $\Omega_2 = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$, muni de la probabilité uniforme $\mathbb{P}_2$. Pour chaque réalisation $\omega = (\omega_1, \omega_2)$, ω_1 correspond au résultat du premier dé et ω_2 au second. On a :

$$\mathbb{P}_2(\text{la somme est 7}) = \frac{\text{Card } \{(\omega_1, \omega_2) \in \Omega_2; \omega_1 + \omega_2 = 7\}}{\text{Card } \Omega_2} = \frac{6}{36} = \frac{1}{6}\,.$$

En revanche on a $\mathbb{P}_2(\text{la somme est 12}) = 1/36$. On peut alors calculer la probabilité des évènements $\{\text{la somme est } x\}$, pour $x \in \{2, \ldots, 12\}$. On détermine ainsi la probabilité sur Ω_1 qui modélise l'expérience. Ce n'est pas la probabilité uniforme. Cette modélisation est confirmée par l'estimation empirique des probabilités obtenue en itérant de nombreuses fois cette expérience.

Dans le choix de Ω_2, on a distingué les dés : premier et second dé. Si on ne distingue pas les deux dés (on dit que les dés sont indiscernables), il est naturel de choisir pour espace d'états l'espace des paires $\Omega_3 = \{\omega = \{\omega_1, \omega_2\}; 1 \leq \omega_1 \leq \omega_2 \leq 6\}$, muni de la probabilité uniforme. On remarque que Card $\Omega_3 = 21$. On obtient alors $\mathbb{P}_3(\text{la somme est 7}) = 3/21 = 1/7$ et $\mathbb{P}_3(\text{la somme est 12}) = 1/21$. Ces valeurs sont très différentes de celles calculées sous $\mathbb{P}_2$. Pour se convaincre que les dés sont discernables (et que le modèle $(\Omega_3, \mathcal{P}(\Omega_3), \mathbb{P}_3)$ n'est pas adapté), on fait le raisonnement suivant : Lancer deux dés l'un après l'autre ou lancer deux dés de couleurs différentes donne la même modélisation $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$. Enfin le résultat ne change pas si l'on jette deux dés simultanément qu'ils soient de même couleur ou non. Donc les dés sont discernables. La modélisation est bien donnée par $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$.

Le problème des objets indiscernables apparaît de manière cruciale en mécanique quantique. Si les particules physiques élémentaires appelées fermions (par exemple les électrons) ne peuvent occuper le même état quantique, il n'en est pas

de même pour les bosons (par exemple les photons). Par exemple les spins des photons prennent leurs valeurs dans $\{-1, 1\}$. Ils sont distribués suivant la probabilité uniforme sur les configurations. Tout se passe comme si on ne pouvait pas distinguer les particules les unes des autres.

I.5 Dénombrement

Dans le cas où $\mathbb{P}$ est la probabilité uniforme, sur un ensemble fini Ω, pour déterminer la probabilité d'un évènement A, il suffit de calculer son cardinal (cf. (I.3)). On est ainsi ramené à un problème de dénombrement.

On rappelle quelques notions élémentaires de dénombrement.

1. Le nombre de **permutations** de $\{1, \dots, n\}$ (ou de bijections d'un ensemble à n éléments dans un ensemble à n éléments) est $n! = n(n-1)\cdots 1$. Par convention, on pose $0! = 1$.
2. Le nombre d'**arrangements** de k éléments dans un ensemble à n éléments avec $k \leq n$ (ou d'injections de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$) est $A_n^k = \dfrac{n!}{(n-k)!}$.
3. Le nombre de sous ensembles à k éléments dans un ensemble à n éléments avec $k \leq n$, est le coefficient binomial $C_n^k = \dfrac{n!}{k!(n-k)!}$, noté aussi $\binom{n}{k}$. On rappelle également la **formule du binôme** : $(x+y)^n = \displaystyle\sum_{k=0}^{n} C_n^k \, x^k y^{n-k}$. En particulier on a $\sum_{k=0}^{n} C_n^k = 2^n$.

Exercice I.2.
On considère une classe de n élèves. On suppose que toutes les années ont 365 jours (i.e. il n'y a pas d'année bissextile).

1. Quelle est la probabilité, p_n, pour que deux élèves au moins aient la même date d'anniversaire ? Trouver le plus petit entier n_1 tel que $p_{n_1} \geq 0.5$. Calculer p_{366}.
2. Quelle est la probabilité, q_n, pour qu'au moins un élève ait la même date d'anniversaire que Socrate ? Calculer q_{n_1} et q_{366}.

△

Correction I.2. Pour répondre à la première question on définit d'abord l'espace de probabilités : $\Omega = \{1, \dots, 365\}^n$ avec $\omega = (\omega_1, \dots, \omega_n)$ où ω_i est la date d'anniversaire de l'élève i. On choisit la probabilité uniforme sur Ω. On a alors :

$$
\begin{aligned}
p_n &= \mathbb{P}(\text{au moins 2 élèves ont la même date d'anniversaire}) \\
&= 1 - \mathbb{P}(\text{tous les élèves ont des dates d'anniversaire différentes}) \\
&= 1 - \mathbb{P}(\{\omega; \omega_i \neq \omega_j, \forall i \neq j\}) \\
&= 1 - \frac{\text{Card}\,\{\omega; \omega_i \neq \omega_j, \forall i \neq j\}}{365^n} \\
&= 1 - \frac{\text{Card}\,\{\text{injections de } \{1,\dots,n\} \text{ dans } \{1,\dots,365\}\}}{365^n} \\
&= \begin{cases} 1 - \dfrac{365!}{(365-n)!365^n} & \text{si } n \leq 365, \\ 1 & \text{si } n \geq 366. \end{cases}
\end{aligned}
$$

On obtient les valeurs numériques suivantes :

$$p_{22} \simeq 0.476; \qquad p_{23} \simeq 0.507; \qquad p_{366} = 1.$$

(Pour les petites valeurs de n, on a l'approximation suivante :

$$
\begin{aligned}
\frac{365!}{(365-n)!365^n} = \prod_{k=1}^{n-1}\left(1 - \frac{k}{365}\right) &= \mathrm{e}^{\sum_{k=1}^{n-1}\log(1-\frac{k}{365})} \\
&\simeq \mathrm{e}^{-\sum_{k=1}^{n-1}\frac{k}{365}} = \mathrm{e}^{-n(n-1)/730} \simeq \mathrm{e}^{-n^2/730}.
\end{aligned}
$$

On obtient $p_n \simeq 1/2$ pour $\mathrm{e}^{-n^2/730} \simeq 1/2$ soit $n \simeq \sqrt{730\log(2)}$. Comme $\log(2) \simeq 0.7$, il vient $n \simeq \sqrt{511}$ soit $n \in \{22, 23\}$.) En fait, les naissances ne sont pas uniformément réparties sur l'année. Les valeurs statistiques de p_n sont donc plus élevées.

Pour la deuxième question, on a, en notant x la date d'anniversaire de Socrate :

$$
\begin{aligned}
q_n &= \mathbb{P}(\text{au moins un élève a son anniversaire le jour } x) \\
&= 1 - \mathbb{P}(\text{tous les élèves ont leur date d'anniversaire différente de } x) \\
&= 1 - \mathbb{P}(\{\omega; \omega_i \neq x, \forall i \in \{1,\dots,n\}\}) \\
&= 1 - \frac{\text{Card}\,\{\omega; \omega_i \neq x, \forall i \in \{1,\dots,n\}\}}{365^n} \\
&= 1 - \left(\frac{364}{365}\right)^n.
\end{aligned}
$$

On obtient les valeurs numériques suivantes :

$$q_{23} \simeq 0.061; \qquad q_{366} \simeq 0.634.$$

Les valeurs p_n et q_n sont très différentes. ▲

De manière générale, dans les problèmes de dénombrement, il faut donner beaucoup d'attention à la rédaction et à la lecture de l'énoncé. Il n'est pas rare de voir des personnes confirmées commettre des erreurs. Pour comprendre le problème, on peut essayer de le simuler par ordinateur. La programmation permet parfois de clarifier les idées.

I.6 Probabilités conditionnelles

On considère le lancer de deux dés équilibrés. On désire connaître la probabilité de $A = \{\text{la somme des deux dés est plus grande que } 10\}$ sachant $B = \{\text{la face du deuxième dé est cinq}\}$. Au lieu de tenir compte de toutes les réalisations on ne garde que celles où le deuxième dé indique 5. La fréquence empirique est donc $\dfrac{N(A\cap B)}{N(B)} = \dfrac{N(A\cap B)}{N}\dfrac{N}{N(B)}$. On trouve environ $\dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. Cela conduit à la définition suivante.

Définition I.6. *Soit A et B deux évènements tels que $\mathbb{P}(B) > 0$. La probabilité **conditionnelle** de A **sachant** B notée $\mathbb{P}(A|B)$ est définie par :*

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.}$$

Proposition I.7. *Soit A et B deux évènements tels que $\mathbb{P}(B) > 0$. On a :*

$$\mathbb{P}(A\cap B) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Si de plus $\mathbb{P}(B^c) > 0$ (et donc $\mathbb{P}(B) < 1$), on a la formule de décomposition suivante :

$$\boxed{\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c),}$$

et la formule de Bayes :

$$\boxed{\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.}$$

La démonstration de la proposition est immédiate à partir de la définition et de l'additivité de $\mathbb{P}$. En utilisant la formule de décomposition (I.2), les deux dernières formules ci-dessus se généralisent de la manière suivante.

Proposition I.8. *Soit $(B_i, i \in I)$ une collection finie ou dénombrable d'évènements disjoints deux à deux tels que $\sum_{i\in I} \mathbb{P}(B_i) = 1$ et $\mathbb{P}(B_i) > 0$ pour tout $i \in I$. Alors pour tout évènement A, on a :*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i),$$

et, pour tout $i_0 \in I$,

$$\mathbb{P}(B_{i_0}|A) = \frac{\mathbb{P}(A|B_{i_0})\mathbb{P}(B_{i_0})}{\sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Exercice I.3.
On suppose que l'on a autant de chance d'avoir une fille ou un garçon à la naissance. Votre voisin de palier vous dit qu'il a deux enfants.

1. Quelle est la probabilité qu'il ait au moins un garçon ?
2. Quelle est la probabilité qu'il ait un garçon, sachant que l'aînée est une fille ?
3. Quelle est la probabilité qu'il ait un garçon, sachant qu'il a au moins une fille ?
4. Vous téléphonez à votre voisin. Une fille décroche le téléphone. Vous savez que dans les familles avec un garçon et une fille, la fille décroche le téléphone avec probabilité p, quelle est la probabilité que votre voisin ait un garçon ?
5. Vous sonnez à la porte de votre voisin. Une fille ouvre la porte. Sachant que l'aîné(e) ouvre la porte avec probabilité p, et ce indépendamment de la répartition de la famille, quelle est la probabilité que votre voisin ait un garçon ?

△

I.7 Indépendance

Lorsqu'on lance un dé plusieurs fois, les différents résultats ne dépendent pas les uns des autres ; ils sont indépendants. La définition suivante donne un cadre mathématique général à la notion d'indépendance entre évènements.

Définition I.9. *On dit que deux évènements A et B sont **indépendants** si :*

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).}$$

Les évènements $(A_i, i \in I)$, où I est une famille quelconque d'indices, sont indépendants si, pour toute famille finie d'indices $J \subseteq I$, on a :

$$\boxed{\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i).}$$

La remarque suivante est une conséquence directe de la définition.

Remarque I.10. Soit $(A_i, i \in I)$ une famille au plus dénombrable d'évènements indépendants. Si $J \subset I$, alors la famille $(A_j, j \in J)$ est une famille (au plus dénombrable) d'évènements indépendants. ◊

Remarque I.11. Si A, B sont des évènements indépendants avec $\mathbb{P}(B) > 0$, alors on déduit de la définition ci-dessus que : $\mathbb{P}(A|B) = \mathbb{P}(A)$. ◊

Exemple. Si on choisit au hasard une carte dans un jeu de 52 cartes, alors sa couleur (pique, cœur, carreau, trèfle) est indépendante de sa valeur (as, roi,...). On a par exemple $\mathbb{P}$(la carte est un as) $= 1/13$ et $\mathbb{P}$(la carte est un pique) $= 1/4$. De plus la probabilité pour que la carte soit un as de pique est égale à $1/52$. C'est le produit des deux probabilités $\mathbb{P}$(la carte est un as) et $\mathbb{P}$(la carte est un pique). On en déduit donc l'indépendance des deux évènements {la carte est un as} et {la carte est un pique}. ◊

I.8 Modélisation (II)

On désire modéliser le lancer de deux dés. Pour le lancer du premier, le choix naturel de l'espace d'états est l'ensemble des résultats possibles $\Omega_1 = \{1, \ldots, 6\}$ et $\mathbb{P}_1$ est la probabilité uniforme sur Ω_1. Le modèle pour le lancer du premier dé est donc $(\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1)$. De même le modèle pour le lancer du deuxième dé est $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$, où $\Omega_2 = \{1, \ldots, 6\}$ et $\mathbb{P}_2$ est la probabilité uniforme sur Ω_2. L'espace d'états associé au lancer des deux dés est l'**espace produit** :

$$\Omega = \Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2); \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

La probabilité sur Ω est définie de façon unique par la donnée de $\mathbb{P}(\{\omega\}) = \mathbb{P}(\{(\omega_1, \omega_2)\})$. Comme les deux lancers sont indépendants, il est naturel d'imposer que :

$$\mathbb{P}(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}).$$

On dit que $\mathbb{P}$ est la probabilité **produit**. On la note $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$. Le modèle du lancer des deux dés est donc $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On parle d'espace probabilisé produit.

De manière plus générale, on modélise des expériences **indépendantes** à l'aide d'espaces probabilisés produits. Dans le cas où l'on a un nombre fini d'expériences indépendantes et où chacune est modélisée par un espace d'états au plus dénombrable, on peut étendre sans difficulté la construction précédente. Des notions supplémentaires de la théorie de la mesure sont nécessaires lorsque l'on sort de ce cadre.

I.9 Rappels sur les ensembles

On dit qu'un ensemble Ω est **fini** s'il existe $n \in \mathbb{N}^*$ et une bijection de Ω dans $\{1, \dots, n\}$. Le cardinal de Ω est Card $\Omega = n$. On dit que Ω est **dénombrable** s'il existe une bijection de Ω dans $\mathbb{N}$, l'ensemble des entiers naturels. (En fait il suffit de montrer qu'il existe une injection et une surjection de Ω vers $\mathbb{N}$.) Par exemple, l'ensemble des entiers relatifs $\mathbb{Z}$ ainsi que l'ensemble des rationnels $\mathbb{Q}$ sont dénombrables. En revanche l'ensemble des réels $\mathbb{R}$ n'est pas dénombrable. Un ensemble est au plus dénombrable s'il est fini ou dénombrable.

On dit que $(A_i, i \in I)$ forme une partition de A si pour tout i, $A_i \neq \emptyset$, pour tout $i \neq i'$, $A_i \cap A_{i'} = \emptyset$ et $\bigcup_{i\in I} A_i = A$.

On rappelle quelques relations usuelles sur les ensembles. Soit I un ensemble d'indices, $(A_i, i \in I)$ une famille de sous-ensembles de Ω et $B \subset \Omega$. On a :

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c, \quad \text{et} \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c, \tag{I.4}$$

$$B \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (B \cap A_i), \quad \text{et} \quad B \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} (B \cup A_i). \tag{I.5}$$

I.10 Compléments sur les espaces mesurables et les fonctions mesurables

On donne brièvement les principales définitions et propriétés des espaces mesurables et des fonctions mesurables. On renvoie par exemple au livre [12] pour plus d'information sur la théorie de la mesure.

Définition I.12. *Soit $\mathcal{F}$ un ensemble de parties de Ω. On dit que $\mathcal{F}$ est une* ***tribu*** *si :*

1. *$\Omega \in \mathcal{F}$.*
2. *Si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$.*
3. *Si $(A_n, n \in \mathbb{N})$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

Le couple $(\Omega, \mathcal{F})$ est appelé ***espace mesurable****. Un ensemble $A \subset \Omega$ est dit* ***mesurable*** *si $A \in \mathcal{F}$.*

Par exemple l'ensemble des parties de Ω : $\mathcal{P}(\Omega)$ est une tribu. L'ensemble $\{\emptyset, \Omega\}$ est également une tribu. C'est la tribu triviale sur Ω.

Si Ω est au plus dénombrable, la tribu la plus souvent considérée est $\mathcal{P}(\Omega)$. Pour des raisons techniques difficiles mais fondamentales, ce n'est quasiment plus le cas si Ω n'est plus fini ou dénombrable.

Exemple. On considère l'ensemble $\mathbb{R}^d$. La tribu **borélienne** est la plus petite tribu qui contienne tous les ouverts de $\mathbb{R}^d$. On la note $\mathcal{B}(\mathbb{R}^d)$. Tout ensemble appartenant à $\mathcal{B}(\mathbb{R}^d)$ est appelé borélien. Si K est un borélien, on note $\mathcal{B}(K) = \{A \subset K; A \in \mathcal{B}(\mathbb{R}^d)\}$ la tribu borélienne sur K. La mesure de Lebesgue $\lambda(E)$ d'un ensemble E est bien définie si E est un borélien.

On note également $\mathcal{B}(\overline{\mathbb{R}})$ la plus petite tribu qui contienne tous les ouverts de $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. ◊

Exercice I.4.
Montrer que la tribu borélienne existe. Pour cela on remarquera qu'il suffit de vérifier que l'intersection d'une famille de tribus est une tribu. △

Remarque I.13. La tribu borélienne sur $[0,1[$ est différente de $\mathcal{P}([0,1[)$ (voir aussi [14], théorème 2.22 où [16] corollaire 5.2.6 et la remarque 4). La construction du contre-exemple suivant est due à Vitali. On définit la relation d'équivalence : $x \in [0,1[$ et $y \in [0,1[$ sont en relation si $x - y \in \mathbb{Q}$. Soit H un sous ensemble de $[0,1[$ contenant un représentant de chaque classe d'équivalence (H est construit à l'aide de l'axiome du choix). On considère ensuite les ensembles $H_r = \{x + r \mod 1; x \in H\}$, pour $r \in [0,1[\cap\mathbb{Q}$. Cette famille d'ensembles est une partition de $[0,1[$. En particulier $\bigcup_{r\in[0,1[\cap\mathbb{Q}} H_r = [0,1[$ et les ensembles H_r sont disjoints deux à deux. Si $H \in \mathcal{B}([0,1[)$, alors on aurait $\lambda([0,1[) = \sum_{r\in[0,1[\cap\mathbb{Q}} \lambda(H_r)$, où λ est la mesure de Lebesgue. Comme la mesure de Lebesgue est invariante par translation, on a aussi $\lambda(H_r) = \lambda(H)$, ce qui contredit l'égalité précédente. En conclusion, H n'est pas un ensemble borélien. ◊

Définition I.14. *Soit $n \geq 1$. Soit $((\Omega_i, \mathcal{F}_i), 1 \leq i \leq n)$ une famille d'espaces mesurables. L'espace produit est $\Omega = \Omega_1 \times \cdots \times \Omega_n$ et la tribu produit, $\mathcal{F}$, est la plus petite tribu de Ω qui contienne $\{A_1 \times \cdots \times A_n; A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n\}$. L'espace mesurable $(\Omega, \mathcal{F})$ est appelé espace mesurable produit.*

Définition I.15. *Une fonction X de Ω dans Ω' est une application mesurable de $(\Omega, \mathcal{F})$ dans $(\Omega', \mathcal{F}')$ si pour tout $A \in \mathcal{F}'$, on a $X^{-1}(A) \in \mathcal{F}$.*

En pratique, toutes les fonctions que l'on considérera seront mesurables. Il est particulièrement difficile de construire des ensembles ou des fonctions non mesurables. Le résultat suivant découle de la définition.

Proposition I.16. *La composée de deux applications mesurables est mesurable.*

Enfin on donne deux résultats importants sur les fonctions mesurables à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Proposition I.17. *Une application X définie sur Ω à valeurs dans $\mathbb{R}^d$ est une application mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ si et seulement si l'une au moins des conditions suivantes est satisfaite :*

1. *Pour tout ouvert $O \subset \mathbb{R}^d$, $X^{-1}(O) \in \mathcal{F}$.*
2. *Pour tout $a_1 < b_1, \ldots, a_d < b_d$, on a $X^{-1}([a_1, b_1] \times \cdots \times [a_d, b_d]) \in \mathcal{F}$.*

En particulier, on a les résultats suivants.

Corollaire I.18. *Une application continue de $\mathbb{R}^d$ dans $\mathbb{R}^l$ est mesurable par rapport aux tribus boréliennes.*

Exemple. Les applications $\varphi_1 : (x, y) \mapsto x + y$ et $\varphi_2 : (x, y) \mapsto xy$ sont continues. Si f et g sont deux applications mesurables de $\mathbb{R}^d$ dans $\mathbb{R}^l$, alors grâce à la proposition I.16, $f + g = \varphi_1(f, g)$ et $fg = \varphi_2(f, g)$ sont des applications mesurables. ◊

On rappelle que si $(a_n, n \in \mathbb{N})$ est une suite de réels, alors la limite supérieure notée $\limsup_{n\to\infty} a_n$ est définie par :

$$\limsup_{n\to\infty} a_n = \inf\{x \in \mathbb{R}; \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \geq n_0, a_n < x\} = \lim_{n\to\infty} \sup_{m\geq n} a_m,$$

avec la convention que $\inf \emptyset = +\infty$. On définit également la limite inférieure par :

$$\liminf_{n\to\infty} a_n = \sup\{x \in \mathbb{R}; \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \geq n_0, a_n > x\} = \lim_{n\to\infty} \inf_{m\geq n} a_m,$$

avec la convention que $\sup \emptyset = -\infty$. Enfin si la limite inférieure et la limite supérieure coïncident, alors on note $\lim_{n\to\infty} a_n$ la limite commune. On dit de plus que la suite $(a_n, n \in \mathbb{N})$ converge.

On admet la proposition suivante.

Proposition I.19. *Soit $(X_n, n \in \mathbb{N})$ une suite de fonctions mesurables de $(\Omega, \mathcal{F})$ dans l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Les fonctions définies sur Ω à valeurs dans l'espace $\overline{\mathbb{R}}$:*

$$\overline{X}(\omega) = \limsup_{n\to\infty} X_n(\omega) \quad et \quad \underline{X}(\omega) = \liminf_{n\to\infty} X_n(\omega)$$

sont mesurables de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. En particulier si la suite de fonctions converge, alors la limite est mesurable.

I.11 Résumé

On rappelle les propriétés suivantes des probabilités.

- σ-additivité. Soit $(A_i, i \in I)$ une suite au plus dénombrable d'évènements disjoints deux à deux, alors on a $\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbb{P}(A_i)$.
- Convergence monotone. Soit $(A_n, n \in \mathbb{N})$ une suite croissante d'évènements, alors on a $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$.
- Formule de décomposition. Soit $(A_i, i \in I)$ une suite au plus dénombrable d'évènements disjoints deux à deux, telle que $\sum_{i\in I} \mathbb{P}(A_i) = 1$, alors pour tout évènement B, on a $\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \cap A_i)$.
- Probabilité conditionnelle. $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ si $\mathbb{P}(B) > 0$.
- Formule de Bayes. $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}$ si $1 > \mathbb{P}(B) > 0$.
- Indépendance. Deux évènements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- La formule de Bayes (resp. la définition d'évènements indépendants) possède une extension pour une famille au plus dénombrable (resp. quelconque) d'évènements.

I.12 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice I.1 p. 8, *Exercice* I.3 p. 13,
Exercice I.2 p. 10, *Exercice* I.4 p. 16.

Exercice I.5.
Montrer que si A et B sont des évènements indépendants, alors A^c et B ainsi que A^c et B^c sont indépendants. △

Exercice I.6.
Soit $A_1, \cdots, A_{n+1}$ des évènements. Montrer que si $\mathbb{P}(A_1 \cap \cdots \cap A_n) > 0$, alors on a la formule dite des probabilités composées :

$$\mathbb{P}(A_1 \cap \cdots \cap A_{n+1}) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_{n+1} \mid A_1 \cap \cdots \cap A_n).$$

△

Exercice I.7.
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. Pour tout évènement $A \in \mathcal{F}$, on note $Q(A) = \mathbb{P}(A \mid B)$. Montrer que Q est une probabilité sur $(\Omega, \mathcal{F})$. △

Exercice I.8.
La **formule du crible**. Soit $A_1, \cdots, A_n$ des évènements.

1. Montrer que $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$.
2. Montrer la formule du crible par récurrence :

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{p=1}^{n} (-1)^{p+1} \sum_{1 \le i_1 < \cdots < i_p \le n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_p}). \tag{I.6}$$

3. Inégalités de Bonferroni. Montrer, par récurrence sur n, que : pour $1 \le m \le n$,

$$\sum_{p=1}^{m} (-1)^{p+1} \sum_{1 \le i_1 < \cdots < i_p \le n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{j_p})$$

est une majoration (resp. minoration) de $\mathbb{P}(\bigcup_{i=1}^{n} A_i)$ lorsque m est impair (resp. pair).

△

Exercice I.9.
Le joueur A possède deux dés à six faces, et le joueur B possède un dé à douze faces. Le joueur qui fait le plus grand score remporte la mise (match nul si égalité). Le jeu est-il équilibré ? On calculera la probabilité que A gagne et la probabilité d'avoir un match nul.

△

Exercice I.10.
Une urne contient r boules rouges et b boules bleues.

1. On tire **avec** remise $p \in \mathbb{N}^*$ boules. Calculer la probabilité pour qu'il y ait p_r boules rouges et p_b boules bleues ($p_r + p_b = p$).
2. On tire **sans** remise $p \leq r + b$ boules. Calculer la probabilité pour qu'il y ait p_r boules rouges et p_b boules bleues ($p_r \leq r$, $p_b \leq b$ et $p_r + p_b = p$).
3. Calculer, dans les deux cas, les probabilités limites quand $r \to \infty$, $b \to \infty$ et $r/(b+r) \to \theta \in]0, 1[$.

△

Exercice I.11.
Les laboratoires pharmaceutiques indiquent pour chaque test sa sensibilité α, qui est la probabilité que le test soit positif si le sujet est malade, et sa spécificité β, qui est la probabilité que le test soit négatif si le sujet est sain. Sachant qu'en moyenne il y a un malade sur 1000 personnes, calculer la probabilité pour que vous soyez un sujet sain alors que votre test est positif, avec $\alpha = 98\%$ et $\beta = 97\%$. Calculer la probabilité d'être malade alors que le test est négatif. Commentaire.

△

Exercice I.12.
Le gène qui détermine la couleur bleue des yeux est récessif. Pour avoir les yeux bleus, il faut donc avoir le génotype *bb*. Les génotypes *mm* et *bm* donnent des yeux marron. On suppose que les parents transmettent indifféremment un de leurs gènes à leurs enfants. La sœur et la femme d'Adrien ont les yeux bleus, mais ses parents ont les yeux marron.

1. Quelle est la probabilité pour qu'Adrien ait les yeux bleus ?
2. Quelle est la probabilité que le premier enfant d'Adrien ait les yeux bleus sachant qu'Adrien a les yeux marron ?
3. Quelle est la probabilité pour que le deuxième enfant d'Adrien ait les yeux bleus sachant que le premier a les yeux marron ?
4. Comment expliquez-vous la différence des résultats entre les deux dernières questions ? △

Exercice I.13.
La justice enquête sur une relation de parenté entre deux personnes. Pour cela elle confie l'analyse des phénotypes sanguins à deux laboratoires indépendants A et B. Ceux-ci délivrent des résultats corrects avec probabilité α (laboratoire A) et β (laboratoire B). La probabilité pour que deux personnes prises au hasard aient le même phénotype vaut τ. On note I l'évènement : les phénotypes sont identiques, et A_+ (resp. B_+) l'évènement : le laboratoire A (resp. B) assure que les phénotypes sont identiques.

1. Expliquer pourquoi l'indépendance des deux laboratoires implique entre autre que $\mathbb{P}(A_+ \cap B_+ \mid I) = \alpha\beta$.
2. Calculer la probabilité que les deux personnes aient le même phénotype sachant que les résultats des deux laboratoires sont positifs.
3. A.N. $\alpha = \beta = 0,9$ et $\tau = 10^{-3}$. Commentaire.

△

Exercice I.14.
Afin de savoir si les élèves travaillent indépendamment ou en groupe, un enseignant donne m exercices à une classe de n élèves. Chaque élève choisit k exercices parmi les m.

1. Calculer la probabilité pour que les élèves aient tous choisi une combinaison fixée de k exercices.
2. Calculer la probabilité pour que tous les élèves aient choisi les k mêmes exercices.
3. Calculer la probabilité pour qu'une combinaison fixée à l'avance, n'ait pas été choisie.
4. Calculer la probabilité pour qu'il existe au moins une combinaison de k exercices qui n'ait pas été choisie. (On utilisera la formule du crible (I.6) cf. exercice I.8)
5. A.N. Donner les résultats pour $n = 20$, $m = 4$, $k = 2$. Comparer les valeurs pour les questions 1 et 2 puis 3 et 4.

△

Exercice I.15.
Paradoxe de Bertrand (1889). On considère trois cartes : une avec les deux faces rouges, une avec les deux faces blanches, et une avec une face rouge et une face blanche. On tire une carte au hasard. On expose une face au hasard. Elle est rouge. Parieriez-vous que la face cachée est blanche ? Pour vous aider dans votre choix :

1. Déterminer l'espace de probabilité.
2. Calculer la probabilité que la face cachée soit blanche sachant que la face visible est rouge.

△

Exercice I.16.
On utilise dans cet exercice la formule du crible (I.6) (cf exercice I.8).

1. Pour fêter leur réussite à un concours, n étudiants se donnent rendez-vous dans un chalet. En entrant chaque personne dépose sa veste dans un vestiaire. Au petit matin, quand les esprits ne sont plus clairs, chacun prend une veste au hasard. Quelle est la probabilité pour qu'une personne au moins ait sa propre veste ?
2. En déduire le nombre de permutations de $\{1, \ldots, n\}$ sans point fixe (problème formulé par de Montmort en 1708) [1]
3. En s'inspirant de la question 1, calculer la probabilité $\pi_n(k)$ pour que k personnes exactement aient leur propre veste.
4. Calculer la limite $\pi(k)$ de $\pi_n(k)$ quand n tend vers l'infini. Vérifier que la famille $(\pi(k), k \in \mathbb{N})$ détermine une probabilité sur $\mathbb{N}$. Il s'agit en fait de la loi de Poisson.

△

Exercice I.17.
On cherche à ranger n paires de chaussettes identiques dans p tiroirs d'une commode.

1. Calculer le nombre N de configurations possibles. On montrera que N peut s'interpréter comme le nombre de résultats possibles d'un tirage successif de $n + p - 1$ boules (sans remise) dans une urne contenant n boules blanches et $p - 1$ boules noires.
2. Si les configurations sont équiprobables, calculer la probabilité pour que le tiroir du haut contienne une seule paire de chaussettes.
3. On range les paires de chaussettes au hasard. Les paires sont indépendantes et ont la même probabilité d'être rangées dans chacun des p tiroirs. Calculer la probabilité pour que le tiroir du haut contienne une seule paire de chaussettes.
4. Expliquer pourquoi les deux résultats sont différents (on détaillera pour $n = p = 2$).

△

1. Voir L. Takacs (The problem of coincidences, *Arch. Hist. Exact Sci.* 21 :3 (1980), 229-244) pour une étude historique du problème des coïncidences vu par les probabilistes.

Exercice I.18.
On reprend l'exercice I.17.

1. Sous les hypothèses de la question 2 (chaussettes indiscernables), calculer $\pi_{n,p}(k)$, la probabilité pour que le tiroir du haut contienne exactement k chaussettes. Calculer la limite $\pi_\lambda(k)$ quand $n \to \infty$ et $n/p \to \lambda \in]0, \infty[$. Vérifier que π_λ est une probabilité sur $\mathbb{N}$.
2. Même question avec les hypothèses de la question 3 (chaussettes discernables).

△

Exercice I.19.
On utilise dans cet exercice la formule du crible (I.6) (cf exercice I.8). Soit $1 \leq k \leq n$.

1. Calculer à l'aide de la formule du crible le nombre de surjections de $\{1, \cdots, n\}$ dans $\{1, \cdots, k\}$.
2. En déduire $\left\{ {n \atop k} \right\}$, le nombre de partitions d'un ensemble à n éléments en k sous-ensembles non vides. Les nombres $\left\{ {n \atop k} \right\}$ sont appelés les nombres de Stirling de deuxième espèce.
3. Montrer que :

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\}, \quad \left\{ {n \atop 1} \right\} = 1, \quad \left\{ {n \atop k} \right\} = 0 \text{ si } k > n.$$

△

Exercice I.20.
Un prince offre à un prisonnier la grâce suivante. Il dispose de deux urnes identiques et de B boules blanches et N boules noires. Il choisit la répartition des boules dans les urnes, sachant qu'aucune ne peut être vide. Ensuite, il choisit les yeux bandés une urne et une boule dans cette urne. Si la boule est blanche le prisonnier est gracié, sinon il est condamné.

1. Calculer la probabilité $p(b, n)$ que le prisonnier obtienne une boule blanche sachant qu'une urne contient b boules blanches et n noires. (L'autre contient alors $B - b$ boules blanches et $N - n$ noires.)
2. En déduire la répartition que doit choisir le prisonnier.

△

Exercice I.21.
Le problème qui suit est inspiré du jeu télévisé américain "Let's Make a Deal" (1963-1977) présenté par Monty Hall [2]. On considère trois portes : A, B et C.

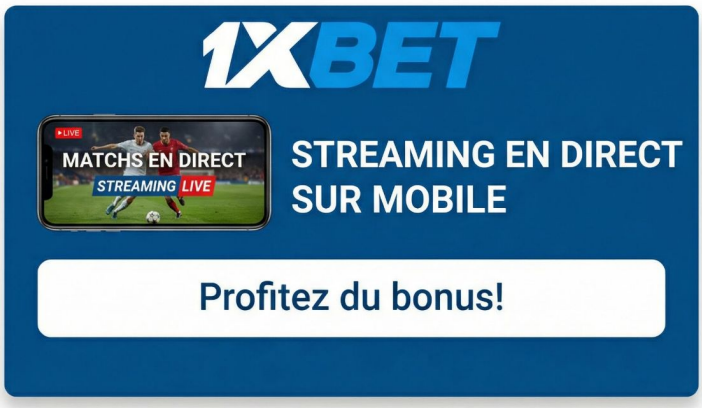
2. Voir le site Wikipedia `http://en.wikipedia.org/wiki/Monty_Hall_problem`.

Derrière l'une d'entre elles se trouve un cadeau et rien derrière les deux autres. Vous choisissez au hasard une des trois portes sans l'ouvrir, par exemple la porte A. À ce moment-là, le présentateur, qui sait derrière quelle porte se trouve le cadeau, ouvre une porte parmi les deux B et C, derrière laquelle il n'y a évidemment rien. On vous propose alors de changer ou non de porte, le but étant d'ouvrir la porte qui cache le cadeau afin de gagner. L'objectif de cet exercice est de déterminer votre meilleure stratégie.

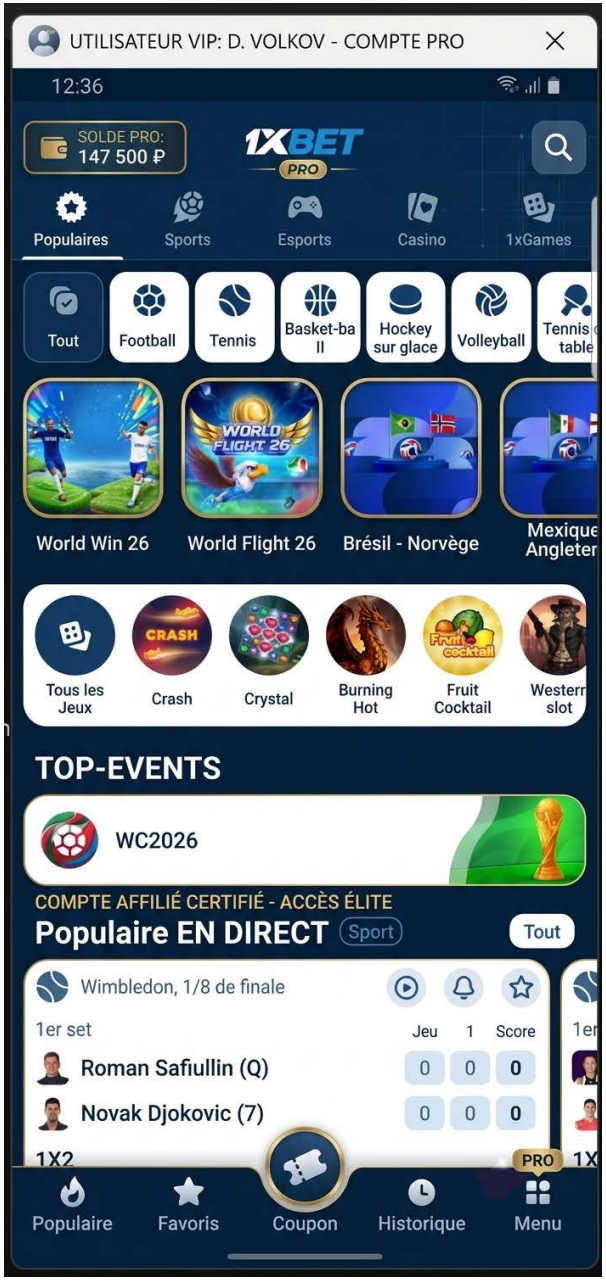
1. On suppose que si le cadeau est derrière la porte A, alors le présentateur choisit au hasard entre les deux autres portes. Calculer la probabilité pour que le cadeau soit derrière la porte B sachant que le présentateur ouvre la porte C. Que faites-vous ?
2. On suppose que si le cadeau est derrière la porte A, alors le présentateur choisit systématiquement la porte B. Que faites-vous si le présentateur ouvre la porte B (respectivement C) ?
3. Montrer que quelle que soit la valeur de la probabilité pour que le présentateur ouvre la porte B (respectivement C) sachant que le cadeau est derrière la porte A, vous avez intérêt à changer de porte. En déduire que la meilleure stratégie consiste à changer systématiquement de porte.
4. Une fois que le présentateur a ouvert une porte, et quel que soit le mécanisme de son choix, vous tirez à pile ou face pour choisir si vous changez ou non de porte. Quelle est votre probabilité de gagner le cadeau ? Vérifier que cette stratégie est moins bonne que la précédente.

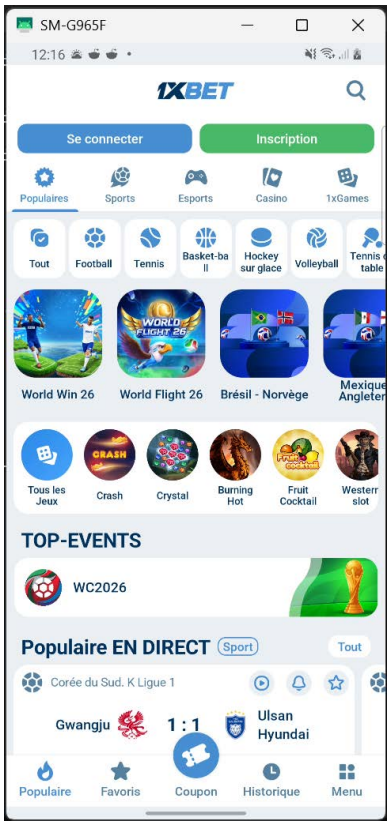
△

ÉTAPES D'INSCRIPTION
pour voir comment creer un compte

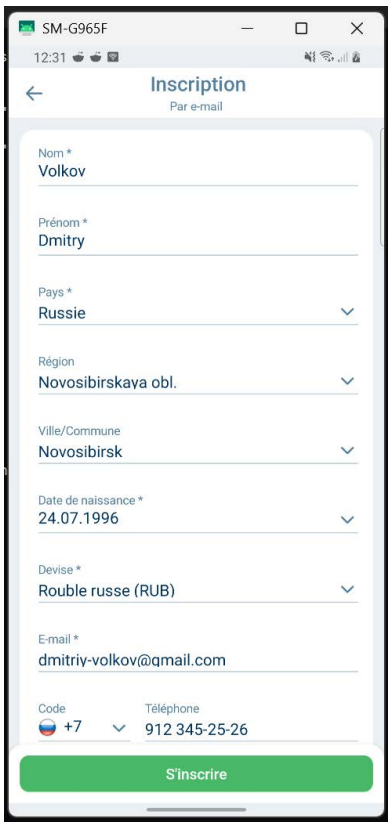

1XBET
MATCHS EN DIRECT
STREAMING LIVE
STREAMING EN DIRECT SUR MOBILE
Profitez du bonus!

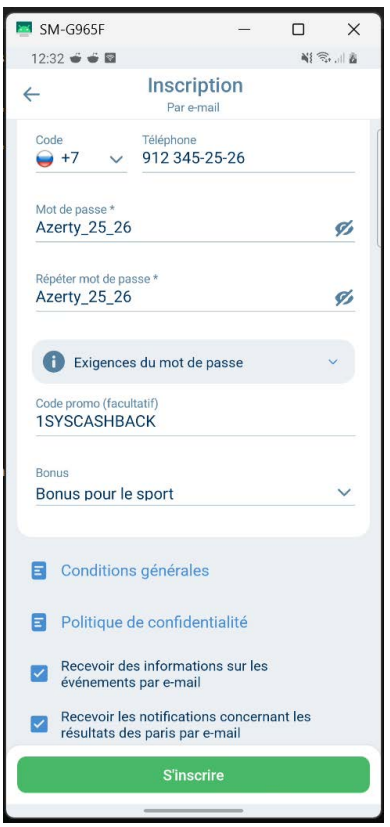
CODE PROMO
1SYSCASHBACK


UTILISATEUR VIP: D. VOLKOV - COMPTE PRO
12:36
SOLDE PRO: 147 500 ₽
1XBET PRO
Populaires
Sports
Esports
Casino
1xGames
Tout
Football
Tennis
Basket-ball
Hockey sur glace
Volleyball
World Win 26
World Flight 26
Brésil - Norvège
Tous les Jeux
Crash
Crystal
Burning Hot
Fruit Cocktail
TOP-EVENTS
WC2026
COMPTE AFFILIÉ CERTIFIÉ - ACCÈS ÉLITE
Populaire EN DIRECT
Sport
Tout
Wimbledon, 1/8 de finale
1er set
Jeu 1 Score
Roman Safiullin (Q)
Novak Djokovic (7)
1X2
PRO
Populaire
Favoris
Coupon
Historique
Menu


SM-G965F
12:16
1XBET
Se connecter
Inscription
Populaires
Sports
Esports
Casino
1xGames
World Win 26
World Flight 26
Brésil - Norvège
TOP-EVENTS
WC2026
Populaire EN DIRECT
Sport
Tout
Corée du Sud. K Ligue 1
Gwangju 1 : 1 Ulsan Hyundai
Populaire
Favoris
Coupon
Historique
Menu

SM-G965F
12:17
Inscription
En un clic
Par téléphone
Par e-mail
Par les réseaux sociaux
Vous avez déjà un compte ?
Se connecter


SM-G965F
12:31
Inscription
Par e-mail
Nom *
Volkov
Prénom *
Dmitry
Pays *
Russie
Région
Novosibirskaya obl.
Ville/Commune
Novosibirsk
Date de naissance *
24.07.1996
Devise *
Rouble russe (RUB)
E-mail *
dmitriy-volkov@gmail.com
Code
+7
Téléphone
912 345-25-26
S'inscrire


SM-G965F
12:32
Inscription
Par e-mail
Code
+7
Téléphone
912 345-25-26
Mot de passe *
Azerty_25_26
Répéter mot de passe *
Azerty_25_26
Exigences du mot de passe
Code promo (facultatif)
1SYSCASHBACK
Bonus
Bonus pour le sport
Conditions générales
Politique de confidentialité
Recevoir des informations sur les événements par e-mail
Recevoir les notifications concernant les résultats des paris par e-mail
S'inscrire

II

Variables aléatoires discrètes

Ce chapitre est consacré à l'introduction des variables aléatoires (à valeurs réelles ou vectorielles) qui permettent de modéliser les résultats d'une expérience en présence d'aléa et à l'étude détaillée des variables aléatoires discrètes, c'est-à-dire des variables aléatoires prenant un nombre au plus dénombrable de valeurs.

La définition des variables aléatoires et de leur loi est donnée au paragraphe II.1. On présente des exemples de variables aléatoires discrètes usuelles aux paragraphes II.2 et II.5. Les paragraphes II.3, II.6 et II.12 sont consacrés aux calculs de loi de variables aléatoires discrètes : calcul des lois des composantes d'une variable aléatoire discrète vectorielle (formule des lois marginales), calcul de la loi pour une transformation d'une variable aléatoire discrète, calcul de lois pour des variables aléatoires à valeurs entières utilisant les fonctions génératrices.

L'espérance d'une variable aléatoire est définie au paragraphe II.7, elle s'interprète comme la moyenne des valeurs possibles de la variable aléatoire pondérées par les probabilités de les observer. On admet les principaux résultats de la théorie de la mesure et de l'intégration. On obtient des formules explicites de calcul d'espérance pour les variables aléatoires discrètes au paragraphe II.8. On définit également la variance d'une variable aléatoire comme l'écart quadratique à la moyenne, et la covariance entre deux variables aléatoires au paragraphe II.9.

On étend la notion d'indépendance aux variables aléatoires discrètes et générales aux paragraphes II.4, II.10 et II.13. Les notions de loi conditionnelle et d'espérance conditionnelle par rapport à une variable aléatoire discrète sont définies au paragraphe II.14.

La loi faible des grands nombres, paragraphe II.11, est un résultat fondamental de la théorie des probabilités. Il stipule que, sous certaines hypothèses, la moyenne arithmétique d'un grand nombre de variables aléatoires (qui est donc une quantité

aléatoire) est proche d'une constante. Ce résultat sera étendu au paragraphe V.4 avec la loi forte des grands nombres.

Enfin, le paragraphe II.15 est consacré à des rappels sur les séries entières qui sont utiles pour les calculs d'espérance de certaines variables aléatoires discrètes.

II.1 Variables aléatoires

On se donne un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition II.1. *On appelle* ***variable aléatoire*** *(v.a.) réelle (resp. vectorielle), toute application X mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou éventuellement dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ (resp. $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Ainsi pour tout borélien A, l'ensemble $\{\omega; X(\omega) \in A\}$ est un évènement. En particulier on a $\{\omega; X(\omega) = x\} \in \mathcal{F}$.

Par convention, on note les variables aléatoires avec des lettres **majuscules** (ex : $X, Y, \ldots$) choisies le plus souvent à la fin de l'alphabet, et les quantités déterministes avec des lettres **minuscules** (ex : $x, y, \ldots$). Afin d'alléger l'écriture, on aura tendance à omettre ω dans les notations. Ainsi si A est un borélien, on notera $\{X \in A\}$ le sous-ensemble de Ω : $\{\omega; X(\omega) \in A\} = X^{-1}(A)$. On écrit $\mathbb{P}(X \in A)$ pour $\mathbb{P}(\{\omega; X(\omega) \in A\})$ et $\mathbb{P}(X = x)$ pour $\mathbb{P}(\{\omega; X(\omega) = x\})$.

La définition laisse une ambiguïté pour les v.a. réelles sur les valeurs $\pm\infty$. On dit que la v.a. réelle X est **finie** si elle prend des valeurs finies p.s. : $\mathbb{P}(X \in \{-\infty, +\infty\}) = 0$.

Si X est une v.a. réelle, on définit une probabilité $\mathrm{P_X}$ sur $\mathbb{R}$ (ou $\overline{\mathbb{R}}$) de la manière suivante : si B est un borélien, alors on pose $\mathrm{P_X(B)} = \mathbb{P}(\mathrm{X} \in \mathrm{B})$. On vérifie que $\mathrm{P_X}$ est bien une probabilité sur l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$). La probabilité $\mathrm{P_X}$ est la **probabilité image** de $\mathbb{P}$ par l'application X. Par définition c'est la **loi** de X. On définit de même la loi d'une v.a. vectorielle en remplaçant $\mathbb{R}$ par $\mathbb{R}^d$ dans ce qui précède. On note aussi la loi de X par $\mathcal{L}(X)$.

Quand on parle de la loi d'une variable aléatoire, on ne fait donc plus référence à l'espace probabilisé sur lequel elle est définie.

On dit que deux v.a. X_1 et X_2 définies respectivement sur $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ sont **égales en loi** si et seulement si elles définissent la même proba-

bilité image. Ceci est équivalent à dire que $\mathbb{P}_1(X_1 \in B) = \mathbb{P}_2(X_2 \in B)$ pour tout évènement B. On le note $\mathcal{L}(X_1) = \mathcal{L}(X_2)$.

On dit que deux v.a. X et Y sont égales presque sûrement si $\mathbb{P}(X = Y) = 1$. On le note $X = Y$ p.s.

En particulier, deux variables aléatoires égales en loi ne sont pas *a priori* définies sur le même espace probabilisé ; et si elles sont définies sur le même espace probabilisé, elles ne sont pas *a priori* p.s. égales. En revanche deux v.a. égales p.s. ont même loi.

Exemple II.2. On a vu au paragraphe précédent comment modéliser le lancer de deux dés à l'aide de l'espace $\Omega = \{1, \dots, 6\}^2$ muni de la probabilité uniforme (et $\mathcal{F} = \mathcal{P}(\Omega)$). Lors d'une réalisation $\omega = (\omega_1, \omega_2) \in \Omega$, ω_1 est le résultat du premier dé, et ω_2 celui du second. La somme des deux dés $S(\omega) = \omega_1 + \omega_2$ définit donc une variable aléatoire. On a par exemple $\mathbb{P}(S = 11) = \mathbb{P}(\{(5, 6), (6, 5)\}) = 2/36 = 1/18$. ◊

II.2 Exemples de variables aléatoires discrètes

On s'intéresse plus particulièrement dans ce paragraphe aux v.a. qui prennent un nombre au plus dénombrable de valeurs.

Soit X une variable aléatoire à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{R}^d$. L'ensemble $\Delta = \{x; \mathbb{P}(X = x) > 0\}$ est la réunion des ensembles $\{x; \mathbb{P}(X = x) \geq 1/n\}$ pour $n \in \mathbb{N}^*$. Comme l'ensemble $\{x; \mathbb{P}(X = x) \geq 1/n\}$ contient au plus n éléments, on en déduit que Δ est au plus dénombrable.

Définition II.3. *Soit X une variable aléatoire à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{R}^d$. On considère l'ensemble (au plus dénombrable) de $\overline{\mathbb{R}}$ ou $\mathbb{R}^d$ défini par $\Delta = \{x; \mathbb{P}(X = x) > 0\}$. On dit que X est une* ***variable aléatoire discrète*** *(v.a.d.) si $\mathbb{P}(X \in \Delta) = 1$. On dit alors que Δ est le* ***support*** *de la loi de la v.a.d. X.*

Proposition II.4. *La loi d'une variable aléatoire discrète réelle ou vectorielle, X, est caractérisée par le support de sa loi Δ et $(\mathbb{P}(X = x), x \in \Delta)$.*

Démonstration. Par σ-additivité, on a : pour tout borélien B,

$$\mathbb{P}(X \in B) = \sum_{x \in \Delta \cap B} \mathbb{P}(X = x) + \mathbb{P}(X \in B \cap \Delta^c).$$

Comme $\mathbb{P}(X \in B \cap \Delta^c) \leq \mathbb{P}(X \in \Delta^c) = 0$, il vient :

$$\mathbb{P}(X \in B) = \sum_{x \in \Delta \cap B} \mathbb{P}(X = x).$$

On en déduit que Δ et $(\mathbb{P}(X = x), x \in \Delta)$ permettent de calculer $\mathbb{P}(X \in B)$ pour tout borélien et donc de déterminer la probabilité image P_X, c'est à dire la loi de X. □

Remarque II.5. Soit X une v.a. telle que $\mathbb{P}(X \in A) = 1$, où l'ensemble A est au plus dénombrable. On pose $\Delta = \{x \in A; \mathbb{P}(X = x) > 0\}$. Par σ-additivité, on a :

$$1 = \sum_{x \in A} \mathbb{P}(X = x) = \sum_{x \in \Delta} \mathbb{P}(X = x) = \mathbb{P}(X \in \Delta).$$

On en déduit que X est une v.a.d. et que Δ est le support de sa loi. ◊

Remarque. Soit X et Y deux v.a. discrètes. Si, pour tout x, on a $\mathbb{P}(X = x) = \mathbb{P}(Y = x)$, alors elles ont même loi.

On remarque que, dans l'exemple II.2, les variables aléatoires S et $14 - S$ sont discrètes et ont même loi, mais qu'elles ne sont pas égales p.s. En effet, on a $\mathbb{P}(S = 14 - S) = \mathbb{P}(S = 7) = 1/6 \neq 1$. ◊

On donne maintenant trois exemples fondamentaux simples.

Exemple II.6. Loi **uniforme** sur $\{1, \ldots, n\}$:

$$\boxed{\mathbb{P}(X = x) = \frac{1}{n}} \quad \forall x \in \{1, \ldots, n\}$$

(et bien sûr $\mathbb{P}(X = x) = 0$ si $x \notin \{1, \ldots, n\}$). Cette loi modélise par exemple le lancer d'un dé équilibré à n faces. ◊

Exemple II.7. Loi de **Bernoulli** de paramètre $p \in [0, 1]$:

$$\boxed{\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p.}$$

La variable aléatoire X prend deux valeurs p.s. 0 ou 1. Cette loi modélise par exemple le lancer d'une pièce biaisée où l'on code 1 pour pile et 0 pour face. ◊

Exemple II.8. **Fonction indicatrice** et loi de Bernoulli. Si A est un évènement, alors on définit la fonction indicatrice :

$$\mathbf{1}_A : \Omega \to \mathbb{R}$$
$$\omega \mapsto \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

Elle prend ses valeurs dans $\{0,1\}$. Il s'agit donc d'une variable aléatoire de loi de Bernoulli. On a bien sûr $\{\mathbf{1}_A = 1\} = A$ et $\{\mathbf{1}_A = 0\} = A^c$ ainsi que les égalités suivantes :

$$\boxed{\mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)} \quad \text{et} \quad \boxed{\mathbb{P}(\mathbf{1}_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)}$$

En particulier, on obtient que $\mathbf{1}_A$ est une variable aléatoire de Bernoulli de paramètre $p = \mathbb{P}(A)$. ◊

II.3 Loi d'un vecteur, lois marginales

Soit X une v.a.d. et B un évènement. On peut décomposer l'évènement B en fonction des valeurs de la v.a.d. X. La famille au plus dénombrable $(\{X = x\}, x \in \Delta)$ forme une collection au plus dénombrable d'évènements disjoints 2 à 2 telle que $\sum_{x\in\Delta} \mathbb{P}(X = x) = 1$. On déduit de (I.2) la formule de décomposition suivante :

$$\boxed{\mathbb{P}(B) = \sum_{x\in\Delta} \mathbb{P}(B, X = x),} \tag{II.1}$$

où par convention $\mathbb{P}(B, X = x) = \mathbb{P}(B \cap \{X = x\})$. Dans cette formule, on peut remplacer la sommation sur Δ par une sommation sur A pourvu que $\Delta \subset A$. Ainsi, si la v.a.d. X est finie, on a $\mathbb{P}(B) = \sum_{x\in\mathbb{R}} \mathbb{P}(B, X = x)$, où tous les termes sont nuls sauf pour $x \in \Delta$.

Soit $d \geq 2$ et $X = (X_1, \ldots, X_d)$ est une v.a.d. à valeurs dans $\mathbb{R}^d$, où l'on note X_i la valeur de la i-ème composante du vecteur X. La formule des lois marginales permet de déterminer la loi de la v.a.d. réelle X_i à partir de la loi du vecteur X. On parle de la **loi marginale** de X_i. Quitte à permuter i et 1, il suffit de calculer la loi de X_1.

Proposition II.9. Formule des lois marginales. *Soit $d \geq 2$. Soit une variable aléatoire discrète vectorielle $X = (X_1, \ldots, X_d)$. La fonction X_1 est une v.a.d. De plus, pour tout $y \in \mathbb{R}$, on a :*

$$\boxed{\mathbb{P}(X_1 = y) = \sum_{x_2\in\mathbb{R},\ldots,x_d\in\mathbb{R}} \mathbb{P}(X_1 = y, X_2 = x_2, \ldots, X_d = x_d).}$$

Dans l'égalité ci-dessus, on remarque que les termes sont positifs et que la somme possède au plus un nombre dénombrable de termes non nuls. Le membre

de droite est donc bien défini. De plus il est non nul pour un nombre au plus dénombrable de valeurs y.

Démonstration. Pour démontrer que X_1 est une v.a.d., on utilise les notions de mesurabilité qui sont détaillées dans le paragraphe I.10. Soit Π_1 la projection de $\mathbb{R}^d$ dans $\mathbb{R}$ qui à $(x_1, \dots, x_d)$ associe x_1. Cette application est continue car l'image réciproque de tout ouvert est un ouvert. Grâce au corollaire I.18, elle est donc mesurable. On a $X_1 = \Pi_1 \circ X$. La fonction X_1 est la composée de deux applications mesurables, elle est donc mesurable. Elle est clairement à valeurs réelles. Pour montrer qu'elle est discrète, il suffit, d'après la remarque II.5, de trouver un ensemble A au plus dénombrable tel que $\mathbb{P}(X_1 \in A) \geq 1$. L'ensemble $A = \Pi(\Delta)$, où Δ est le support de la loi de X, est au plus dénombrable. On a $\{X_1 \in A\} = \{X \in A \times \mathbb{R}^{d-1}\} \supset \{X \in \Delta\}$. Par monotonie, on a $\mathbb{P}(X_1 \in A) \geq 1$. En fait, on peut vérifier que A est le support de la loi de X_1.

La formule des lois marginales se déduit de la formule de décomposition (II.1). On obtient :

$$\begin{aligned}\mathbb{P}(X_1 = y) &= \sum_{(x_1,x_2,\dots,x_n)\in\Delta} \mathbb{P}(X_1 = y, (X_1, X_2, \dots, X_d) = (x_1, x_2, \dots, x_n)) \\ &= \sum_{(x_2,\dots,x_n)\in\mathbb{R}^{d-1}} \mathbb{P}((X_1, X_2, \dots, X_d) = (y, x_2, \dots, x_n)).\end{aligned}$$

□

En fait, tout vecteur formé à partir des coordonnées du vecteur aléatoire X est une v.a.d. vectorielle.

À partir de la loi du couple (X, Y), on peut donc calculer les lois marginales de X et de Y. En revanche, on ne peut déterminer la loi du couple (X, Y) en connaissant seulement les lois marginales de X et de Y, comme le montre l'exemple suivant.

Exemple II.10. On considère le lancer de deux dés, modélisé par l'espace probabilisé $\Omega = \{1, \dots, 6\}^2$ muni de la probabilité uniforme. Si X_1 est la v.a.d. qui représente le résultat du premier dé et X_2 celui du second, on a pour $\omega = (\omega_1, \omega_2) \in \Omega$, $X_1(\omega) = \omega_1$ et $X_2(\omega) = \omega_2$. On a $\mathcal{L}(X_1) = \mathcal{L}(X_2)$, c'est la loi uniforme sur $\{1, \dots, 6\}$. Mais on a $\mathbb{P}(X_1 = 5, X_1 = 6) = 0$ et $\mathbb{P}(X_1 = 5, X_2 = 6) = 1/36$. Ainsi les v.a.d. (X_1, X_2) et (X_1, X_1) n'ont pas même loi alors que toutes les lois marginales sont égales. ◊

II.4 Variables aléatoires discrètes indépendantes (I)

On étend la notion d'indépendance définie pour des évènements au paragraphe I.7 à des variables aléatoires discrètes.

Définition II.11. *Soit $X_1, \dots, X_n$, n v.a.d. (réelles ou vectorielles) définies sur $(\Omega, \mathcal{F}, \mathbb{P})$. Les v.a.d. sont* ***indépendantes*** *si pour tout $x_1, \dots, x_n$, on a :*

$$\boxed{\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).}$$

On peut se restreindre à $x_1 \in A_1, \dots, x_n \in A_n$, où les ensembles A_i sont au plus dénombrables et tels que $\mathbb{P}((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n) = 1$. On remarque que pour montrer que les v.a.d. sont indépendantes, il faut connaître la loi du vecteur.

Les v.a.d. $X_1, \dots, X_n$ sont **indépendantes si et seulement si la loi du n-uplet est le produit des lois marginales**. On le notera $\mathcal{L}(X_1, \dots, X_n) = \mathcal{L}(X_1) \otimes \cdots \otimes \mathcal{L}(X_n)$, et on parlera de loi produit.

Exercice II.1.
Soit A et B deux événements. Montrer que A et B sont indépendants si et seulement si $\mathbf{1}_A$ et $\mathbf{1}_B$ sont des variables aléatoires (discrètes) indépendantes. △

La notion d'indépendance des v.a.d. correspond bien à la notion d'indépendance sur les évènements. Dans l'exemple II.10 du lancer des deux dés, on vérifie facilement que X_1 et X_2 sont indépendants. Cela correspond bien à la notion intuitive d'indépendance entre les deux dés.

La notion de variables aléatoires indépendantes est importante. On en donnera une définition plus générale au paragraphe II.10 et une caractérisation au paragraphe II.13.

Remarque. Indépendance deux à deux. Soit X_1 et X_2 deux v.a.d. définies sur le même espace probabilisé et indépendantes. On suppose qu'elles ont même loi : la loi uniforme sur $\{-1, +1\}$. On note $X_3 = X_1 X_2$. Il est facile de vérifier que X_i et X_j sont indépendantes pour $1 \leq i < j \leq 3$ (on dit que les v.a. X_1, X_2, X_3 sont indépendantes deux à deux), mais que les v.a.d. (X_1, X_2, X_3) ne sont pas indépendantes.

En particulier l'indépendance deux à deux n'implique pas l'indépendance de la famille de v.a.d. ◊

On peut étendre la définition de v.a.d. indépendantes à une famille quelconque de v.a.d.

Définition II.12. *Soit $(X_i, i \in I)$ une famille quelconque de v.a.d. On dit que les v.a.d. sont indépendantes si pour toutes les familles **finies** d'indices $J \subset I$, les v.a.d. $(X_i, i \in J)$ sont indépendantes.*

Exercice II.2.
Soit $(A_i, i \in I)$ une famille quelconque d'évènements. Montrer que les v.a.d. $(\mathbf{1}_{A_i}, i \in I)$ sont indépendantes si et seulement si les évènements $(A_i, i \in I)$ sont indépendants. △

Remarque. Indépendance par paquet. Soit $(X_i, i \in I)$ une famille quelconque de v.a.d. indépendantes. Soit $(I_j, j \in J)$ une partition de I en sous-ensemble finis. On note $Z_j = (X_i, i \in I_j)$. Alors les v.a.d. $(Z_j, j \in J)$ sont indépendantes. ◊

II.5 Schéma de Bernoulli et autres exemples

Avant de donner d'autres exemples de variables aléatoires discrètes, on démontre que l'on peut modéliser un nombre infini de lancers d'une même pièce (éventuellement biaisée).

Théorème II.13. *Il existe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une suite de v.a.d. $(X_n, n \in \mathbb{N}^*)$ indépendantes et de même loi de Bernoulli de paramètre p. Cette famille de v.a.d. est appelée un schéma de Bernoulli.*

Ce théorème découle directement de théorèmes plus généraux d'existence de variables aléatoires (théorème d'extension de Kolmogorov). Toutefois on en donne une démonstration élémentaire, utilisant la mesure de Lebesgue, que l'on pourra omettre dans une première lecture.

Démonstration. On choisit $\Omega = [0, 1]$, la tribu borélienne sur $[0, 1]$ pour $\mathcal{F}$ et la mesure de Lebesgue λ sur $[0, 1]$ pour $\mathbb{P}$. Ainsi pour tout borélien $A \subset [0, 1]$, sa probabilité $\mathbb{P}(A)$ sera sa mesure de Lebesgue $\lambda(A)$. Cela définit bien un espace probabilisé. On construit les fonctions X_n de $[0, 1]$ dans $\{0, 1\}$ en utilisant une décomposition de tout réel de type dyadique. (Pour $p = 1/2$, $X_n(x)$ est le n-ième coefficient de x dans son développement dyadique.) On pose :

$$X_1(x) = 1 \quad \text{si} \quad x \in [0, p[, \quad \text{et} \quad 0 \quad \text{sinon},$$
$$X_2(x) = 1 \quad \text{si} \quad x \in [0, p^2[\cup [p, p + p(1-p)[, \quad \text{et} \quad 0 \quad \text{sinon},$$

et plus généralement, pour tout entier $n \geq 1$,

$$X_n(x) = 1 \quad \text{si} \quad x \in A_n = \bigcup_{k=0}^{2^{n-1}-1} [a_{2k}^n, a_{2k+1}^n[, \quad \text{et} \quad 0 \quad \text{sinon}.$$

La suite $(a_k^n, k \in \{0, \ldots, 2^n\}, n \in \mathbb{N})$ est définie par la récurrence suivante : $a_0^0 = 0, a_1^0 = 1$, et pour tout entier $n \geq 1$, $a_{2^n}^n = 1$, et : pour tout $k \in \{0, \ldots, 2^{n-1} - 1\}$,

$$a_{2k}^n = a_k^{n-1}, \quad \text{et} \quad a_{2k+1}^n = a_{2k}^n + p(a_{k+1}^{n-1} - a_k^{n-1}).$$

Les ensembles A_n sont des boréliens. Comme $X_n = \mathbf{1}_{A_n}$, les fonctions X_n sont donc des v.a.d. Elles suivent des lois de Bernoulli de paramètre p car :

$$\mathbb{P}(X_n = 1) = \lambda(A_n) = \lambda\left(\bigcup_{k=0}^{2^{n-1}-1} [a_{2k}^n, a_{2k+1}^n[\right)$$
$$= \sum_{k=0}^{2^{n-1}-1} p(a_{k+1}^{n-1} - a_k^{n-1}) = p a_{2^{n-1}}^{n-1} = p.$$

De manière plus générale si A s'écrit comme une réunion d'intervalles du type $[a_{k_i}^{n-1}, a_{k_i+1}^{n-1}[$, c'est-à-dire $A = \bigcup_{i \in I} [a_{k_i}^{n-1}, a_{k_i+1}^{n-1}[$, alors on a :

$$\lambda(A \cap A_n) = p\lambda(A). \tag{II.2}$$

En effet, par définition de A_n, on a :

$$A \cap A_n = \bigcup_{i \in I} [a_{k_i}^{n-1}, a_{k_i}^{n-1} + p(a_{k_i+1}^{n-1} - a_{k_i}^{n-1})[$$
$$\text{et} \quad \lambda(A \cap A_n) = \sum_{i \in I} p(a_{k_i+1}^{n-1} - a_{k_i}^{n-1}) = p\lambda(\bigcup_{i \in I} [a_{k_i}^{n-1}, a_{k_i+1}^{n-1}[) = p\lambda(A).$$

On montre que les v.a.d. $(X_n, n \in \mathbb{N}^*)$ sont indépendantes. Grâce à la définition II.12 et à l'exercice II.2, il suffit de montrer que pour toute famille finie $J \subset \mathbb{N}^*$, les évènements $(A_j, j \in J)$ sont indépendants. Il faut donc vérifier que pour tout $n \geq 2$, et $1 \leq j_1 < \ldots < j_n$, on a $\lambda(A_{j_1} \cap \ldots \cap A_{j_n}) = p^n$. Mais il est clair que $A_{j_1} \cap \ldots \cap A_{j_{n-1}}$ s'écrit comme la réunion d'intervalles du type $[a_{k_i}^{j_n-1}, a_{k_i+1}^{j_n-1}[$. On déduit donc de (II.2) que $\lambda(A_{j_1} \cap \ldots \cap A_{j_{n-1}} \cap A_{j_n}) = p\lambda(A_{j_1} \cap \ldots \cap A_{j_{n-1}})$, et donc on a par récurrence $\lambda(A_{j_1} \cap \ldots \cap A_{j_n}) = p^n$. Les v.a.d. $(X_n, n \in \mathbb{N}^*)$ sont donc indépendantes. □

Le schéma de Bernoulli $(X_n, n \in \mathbb{N}^*)$ permet de modéliser un jeu infini de pile ou face avec une pièce biaisée (de paramètre p). La v.a.d. X_n modélise le résultat du n-ième lancer (1 si pile apparaît, 0 sinon). On continue la série d'exemples de v.a.d.

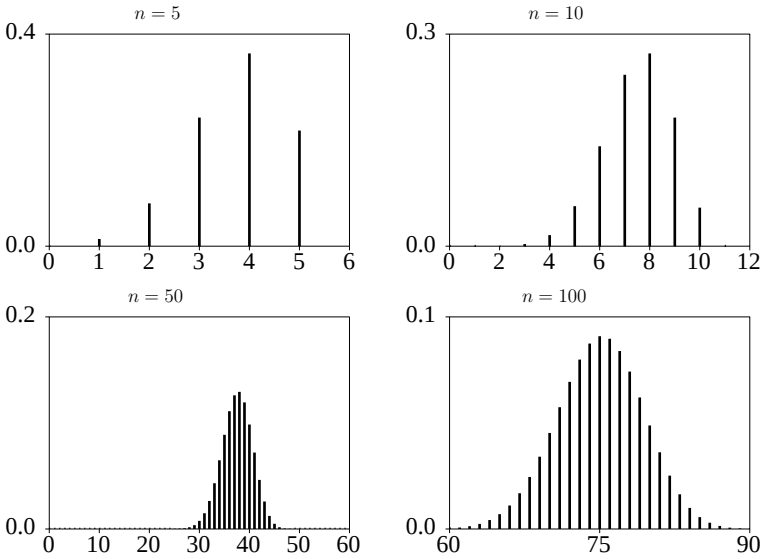


Figure II.1. Loi binomiale de paramètre (n, p), avec $p = 3/4$.

Exemple II.14. Loi **binomiale** $\mathcal{B}(n, p)$ où le paramètre $n \geq 1$ est un entier et $p \in [0, 1]$. Soit S_n, le nombre de fois où pile est apparu lors de n lancers successifs d'une pièce biaisée de paramètre p. Ainsi $S_n = \sum_{i=1}^{n} X_i$, où $(X_i, i \in \mathbb{N}^*)$ est un schéma de Bernoulli. La loi de S_n est par définition la loi $\mathcal{B}(n, p)$. Les figures II.1 et II.2 représentent $\mathbb{P}(S_n = k)$ avec $0 \leq k \leq n$ pour la loi binomiale et différentes valeurs des paramètres. On étudiera par la suite deux limites : n grand à p fixé, et n grand à np fixé.
Pour $k \in \{0, \ldots, n\}$, on a :

$$\begin{aligned}\mathbb{P}(S_n = k) &= \mathbb{P}(X_1 + \cdots + X_n = k) \\ &= \sum_{\substack{x_1 \in \{0,1\}, \ldots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).\end{aligned}$$

En utilisant l'indépendance des v.a.d., il vient

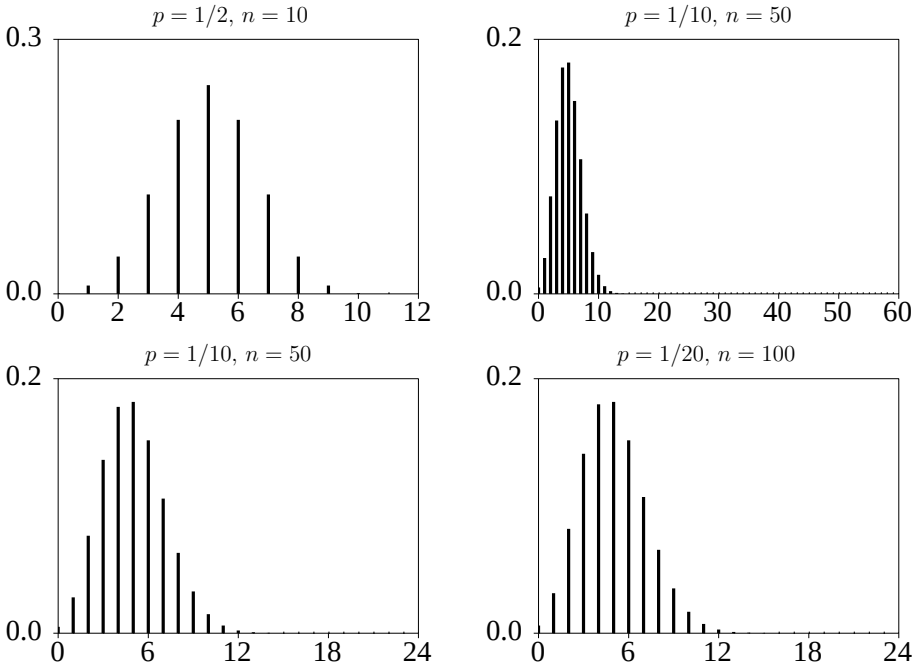


Figure II.2. Loi binomiale de paramètre (n, p), avec $np = 5$.

$$\begin{aligned}\mathbb{P}(S_n = k) &= \sum_{\substack{x_1\in\{0,1\},\dots,x_n\in\{0,1\}\\ x_1+\cdots+x_n=k}} \mathbb{P}(X_1 = x_1)\cdots\mathbb{P}(X_n = x_n)\\ &= \sum_{\substack{x_1\in\{0,1\},\dots,x_n\in\{0,1\}\\ x_1+\cdots+x_n=k}} \prod_{i;\, x_i=1} \mathbb{P}(X_i = 1) \prod_{j;\, x_j=0} \mathbb{P}(X_j = 0)\\ &= \sum_{\substack{x_1\in\{0,1\},\dots,x_n\in\{0,1\}\\ x_1+\cdots+x_n=k}} p^k(1-p)^{n-k}.\end{aligned}$$

Le nombre de parties à k éléments dans un ensemble à n éléments est C_n^k. On en déduit que :

$$\boxed{\mathbb{P}(S_n = k) = C_n^k p^k (1-p)^{n-k} \quad \text{pour } k \in \{0,\dots,n\}.}$$

À l'aide de la formule du binôme, on a $\mathbb{P}(S_n \in \{0,\dots,n\}) = 1$. Le support de la loi binomiale $\mathcal{B}(n,p)$ est donc $\{0,\dots,n\}$ si $p \in]0,1[$. ◊

Exemple II.15. Loi **géométrique** de paramètre $p \in]0,1]$. Soit T, la première fois où l'on obtient un pile dans un jeu infini de pile ou face avec une pièce biaisée de

paramètre p. Ainsi $T = \inf\{n \geq 1; X_n = 1\}$, avec la convention que $\inf \emptyset = +\infty$. La loi de T est par définition la loi géométrique de paramètre p. La figure II.3 représente $\mathbb{P}(T = k)$, $k \geq 0$, pour la loi géométrique et deux valeurs du paramètre.

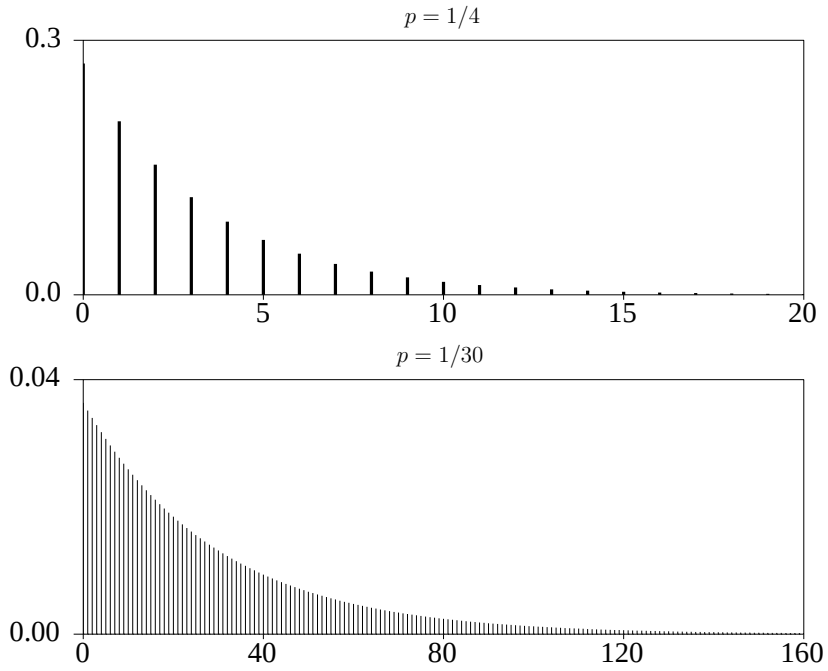


Figure II.3. Loi géométrique de paramètre p.

On remarque que $\{T = 1\} = \{X_1 = 1\}$ et pour $n \geq 2$, $\{T = n\} = \{X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1\}$. On en déduit, en utilisant l'indépendance des v.a.d. $(X_i, i \in \mathbb{N}^*)$ que :

$$\boxed{\mathbb{P}(T = n) = p(1-p)^{n-1}, \quad \text{pour tout} \quad n \in \mathbb{N}^*.}$$

On montre que T est finie p.s. Comme par construction, la v.a. T est à valeurs dans $\mathbb{N} \cup \{+\infty\}$, on déduit de la formule de décomposition que :

$$\mathbb{P}(T < \infty) = \mathbb{P}(T \in \mathbb{N}) = \sum_{n=1}^{\infty} \mathbb{P}(T = n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = 1.$$

La v.a.d. T est donc p.s. finie. ◊

Exemple II.16. Loi de **Poisson** de paramètre $\theta \in [0, \infty)$. Cette famille de loi apparaît comme limite de la loi binomiale de paramètre (n, p) quand $n \to \infty$ et $np \to \theta$ (et donc $p \to 0$) (voir les figures II.2 et II.4). On parle aussi de loi des évènements rares.

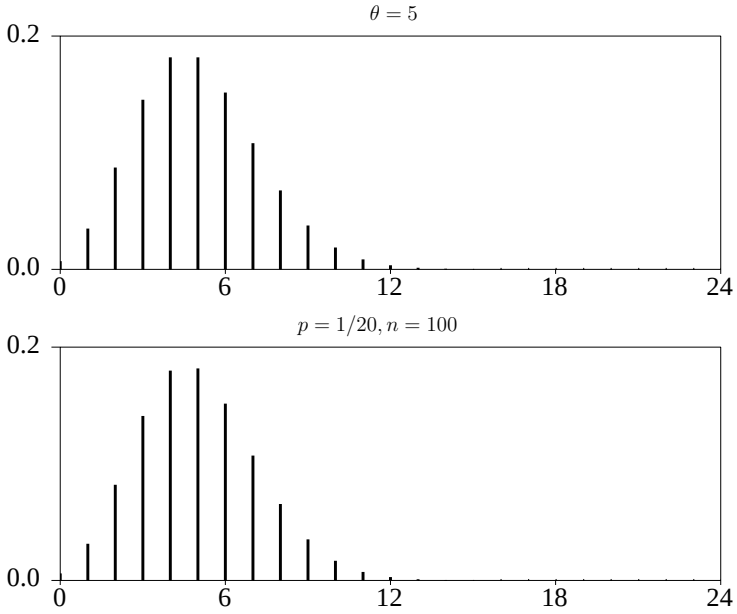


Figure II.4. Loi de Poisson $\mathcal{P}(\theta)$ et loi binomiale $\mathcal{B}(n, p)$, avec $np = \theta$.

En effet si S_n suit une loi $\mathcal{B}(n, p)$, pour $k \in \mathbb{N}$ fixé, et $n > k$, on a :

$$\begin{aligned}
\mathbb{P}(S_n = k) &= C_n^k p^k (1-p)^{n-k} \\
&= \frac{1}{k!}\, p^k n(n-1)\cdots(n-k+1)\, \mathrm{e}^{(n-k)\log(1-p)} \\
&= \frac{1}{k!}\, pn(pn-p)\cdots(pn-pk+p)\, \mathrm{e}^{(pn-pk)\frac{\log(1-p)}{p}} \\
&\to \frac{1}{k!}\, \theta^k\, \mathrm{e}^{-\theta} \quad \text{quand} \quad n \to \infty, \quad p \to 0, \quad \text{et} \quad np \to \theta.
\end{aligned}$$

Par définition une v.a.d. X suit la loi de Poisson de paramètre $\theta \geq 0$ si :

$$\boxed{\mathbb{P}(X = k) = \mathrm{e}^{-\theta}\, \frac{\theta^k}{k!},} \text{ pour tout } k \in \mathbb{N}.$$

La figure II.5 représente la loi de Poisson pour différentes valeurs du paramètre.

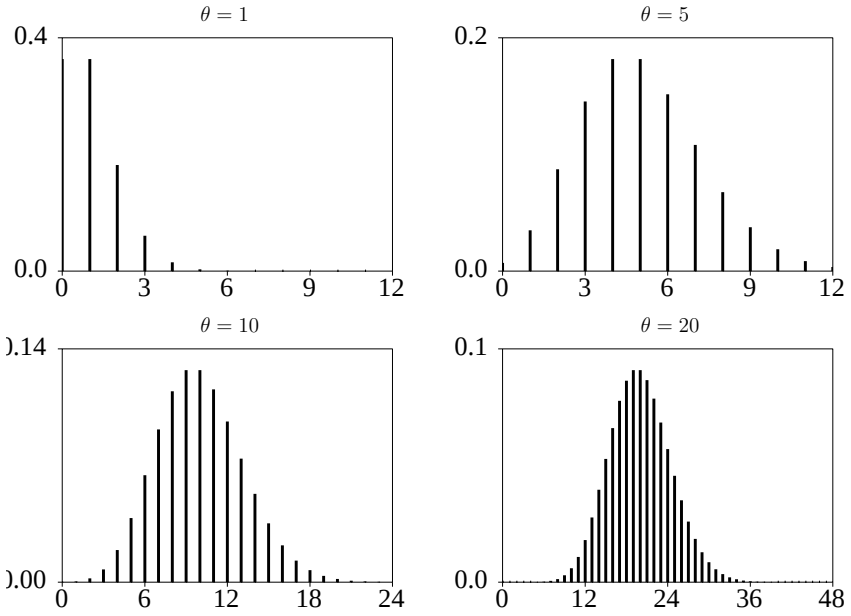


Figure II.5. Loi de Poisson de paramètre θ.

On montre que le support de la loi de Poisson est $\mathbb{N}$. Par σ-additivité, on a :

$$\mathbb{P}(X \in \mathbb{N}) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} \mathrm{e}^{-\theta} \frac{\theta^k}{k!} = 1.$$

En particulier X est finie p.s. ◊

Application. Comment modéliser par exemple le nombre d'appels téléphoniques reçus par une entreprise française en un jour. Une communication téléphonique sur le territoire national a une très faible probabilité d'être pour cette entreprise ($p \approx 0$), en revanche il y a un grand nombre de communications téléphoniques par jour ($n \to \infty$). Si on suppose que les appels sont indépendants, alors le nombre d'appels suit une loi binomiale de paramètre (n, p). Cependant, les paramètres sont tels qu'il est judicieux d'utiliser l'approximation de la loi de Poisson et de modéliser le nombre d'appels téléphoniques par une v.a. de Poisson. Il reste bien sûr à identifier le paramètre θ. ◊

Exercice II.3.
La France a eu 38 médailles dont 13 d'or aux jeux olympiques de Sydney en 2000, sur 928 médailles dont 301 d'or. On estime la population à 6 milliards dont 60

millions en France. Peut-on dire que les sportifs médaillés aux jeux olympiques sont uniformément répartis dans la population mondiale ?

△

II.6 Changement de variable

Soit $\varphi : \mathbb{R}^d \to \mathbb{R}^l$ une fonction mesurable. Soit X une v.a.d. à valeurs dans $\mathbb{R}^d$. On définit la nouvelle fonction $Y = \varphi(X)$. La fonction Y est mesurable (cf. proposition I.16), c'est donc une v.a. Soit Δ le support de la loi de X. On a $\{X \in \Delta\} \subset \{Y \in \varphi(\Delta)\}$. On a donc $\mathbb{P}(Y \in \varphi(\Delta)) = 1$. Comme $\varphi(\Delta)$ est au plus dénombrable, on en déduit que Y est une v.a. discrète. La proposition suivante donne la loi de Y.

Proposition II.17. *La fonction $Y = \varphi(X)$ est une v.a.d. à valeurs dans $\mathbb{R}^l$. Sa loi est donnée par :*

$$\boxed{\mathbb{P}(Y = y) = \sum_{x \in \Delta} \mathbf{1}_{\{\varphi(x)=y\}} \mathbb{P}(X = x)} \quad \textit{pour tout } y \in \mathbb{R}^l.$$

Il s'agit d'une généralisation de la proposition II.9 où $\varphi = \Pi_1$. La démonstration est laissée en exercice.

Exercice II.4.
On considère le lancer de deux dés à 6 faces. Soit $X = (X_1, X_2)$ le couple de v.a.d. représentant le résultat du premier et du second dé. Calculer la loi de la somme des deux faces $S = X_1 + X_2$. Calculer la loi de $\max(X_1, X_2)$ et du vecteur aléatoire $Y = (\max(X_1, X_2), \min(X_1, X_2))$. △

II.7 Espérance d'une variable aléatoire quelconque

La notion d'espérance est la formalisation du concept de "en moyenne". Dans ce paragraphe on énonce la définition de l'espérance pour une v.a. quelconque. On donne des propriétés de l'espérance, que l'on démontrera dans le paragraphe suivant uniquement pour les v.a. discrètes.

Définition II.18. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit X une variable aléatoire réelle positive p.s. (i.e. $\mathbb{P}(X \geq 0) = 1$). On définit son* ***espérance*** $\mathbb{E}[X]$ *comme la limite croissante, éventuellement infinie, suivante :*

$$\mathbb{E}[X] = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbb{P}\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right).$$

Définition II.19. *Soit X est une variable aléatoire réelle. On dit qu'elle est* ***intégrable*** *si $\mathbb{E}[|X|] < \infty$. Dans ce cas-là, on définit son espérance par :*

$$\mathbb{E}[X] = \mathbb{E}[X \mathbf{1}_{\{X \geq 0\}}] - \mathbb{E}[|X| \, \mathbf{1}_{\{X<0\}}].$$

L'espérance d'une variable aléatoire intégrable est finie. Si $X = (X_1, \ldots, X_d)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$, on dit qu'elle est intégrable si les v.a. $X_1, \ldots, X_d$ sont toutes intégrables. L'espérance de X est alors définie par $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$.

Le corollaire suivant est immédiat.

Corollaire II.20. *Deux variables aléatoires intégrables de* ***même loi*** *ont* ***même espérance****.*

Remarque II.21. Soit X une variable aléatoire discrète positive. On a :

$$\frac{k}{2^n}\mathbb{P}(X = x) \leq x\mathbb{P}(X = x) \leq \frac{k+1}{2^n}\mathbb{P}(X = x), \quad x \in [k/2^n, (k+1)/2^n[.$$

En sommant sur $x \in \Delta \cap [k/2^n, (k+1)/2^n[$, il vient :

$$\frac{k}{2^n}\mathbb{P}\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right) \leq \sum_{x \in \Delta \cap [k/2^n, (k+1)/2^n[} x\mathbb{P}(X = x) \leq \frac{k+1}{2^n}\mathbb{P}\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right).$$

En sommant sur k, et en prenant la limite quand n tend vers l'infini, on obtient alors que $\mathbb{E}[X] = \sum_{x\in\Delta} x\mathbb{P}(X = x)$.

En utilisant la formule du changement de variable, proposition II.17, on a, pour une variable aléatoire discrète X quelconque :

$$\mathbb{E}[|X|] = \sum_{x\in\Delta} y\mathbf{1}_{\{|x|=y\}}\mathbb{P}(X = x) = \sum_{x\in\Delta} |x| \, \mathbb{P}(X = x).$$

En conclusion, on obtient que si X est une variable aléatoire discrète intégrable, i.e. si $\sum_{x\in\Delta}|x|\,\mathbb{P}(X=x)<\infty$, alors son espérance est définie par :

$$\boxed{\mathbb{E}[X]=\sum_{x\in\Delta}x\mathbb{P}(X=x),} \tag{II.3}$$

où Δ est le support de la loi de X. ◊

Exemple II.22. Si A est un évènement, la v.a.d. $\mathbf{1}_A$ est positive p.s. et :

$$\boxed{\mathbb{E}[\mathbf{1}_A]=\mathbb{P}(A).}$$

En effet, on a $\mathbb{E}[\mathbf{1}_A]=1.\mathbb{P}(\mathbf{1}_A=1)+0.\mathbb{P}(\mathbf{1}_A=0)=\mathbb{P}(A)$. En particulier, avec $A=\Omega$, on remarque que :

$$\boxed{\mathbb{E}[\mathbf{1}]=1.}$$

◊

Exemple. On considère un dé à 6 faces. Le résultat d'un lancer est une v.a.d. X de loi uniforme sur $\{1,\dots,6\}$. La valeur moyenne d'un lancer de dé est $\mathbb{E}[X]=\sum_{k=1}^{6}k/6=7/2$. ◊

Remarque. Si Ω est au plus dénombrable, alors toutes les variables aléatoires sont discrètes. Si X est intégrable ou positive p.s., on a :

$$\boxed{\mathbb{E}[X]=\sum_{\omega\in\Omega}X(\omega)\mathbb{P}(\{\omega\}).} \tag{II.4}$$

En effet, on a, grâce à la formule de décomposition :

$$\begin{aligned}\mathbb{E}[X]&=\sum_{x\in\Delta}x\mathbb{P}(X=x)\\&=\sum_{x\in\Delta}\sum_{\omega\in\Omega}x\mathbb{P}(\{X=x\}\cap\{\omega\}).\end{aligned}$$

On remarque que si $X(\omega)=x$, alors $\mathbb{P}(\{X=x\}\cap\{\omega\})=\mathbb{P}(\{\omega\})$ et si $X(\omega)\neq x$, alors $\mathbb{P}(\{X=x\}\cap\{\omega\})=\mathbb{P}(\emptyset)=0$. Donc on a

$$\begin{aligned}\mathbb{E}[X]&=\sum_{x\in\Delta}\sum_{\omega\in\Omega}x\mathbf{1}_{\{X(\omega)=x\}}\mathbb{P}(\{\omega\})\\&=\sum_{\omega\in\Omega}X(\omega)\sum_{x\in\Delta}\mathbf{1}_{\{X(\omega)=x\}}\mathbb{P}(\{\omega\})\\&=\sum_{\omega\in\Omega}X(\omega)\mathbb{P}(\{\omega\}).\end{aligned}$$

Les inversions des signes sommes sont justifiées car soit on somme des termes positifs (si X est positive p.s.) soit les sommes sont absolument convergentes (si X est intégrable). ◊

On admet les propriétés suivantes de l'espérance.

Proposition II.23.

1. ***Linéarité***. *Soit X et Y deux v.a. quelconques (réelles ou vectorielles) intégrables, soit $\alpha, \beta \in \mathbb{R}$. Alors la v.a. $\alpha X + \beta Y$ est intégrable, et on a :*
$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$
2. ***Positivité***. *Soit X une v.a. réelle positive p.s., c'est-à-dire telle que $\mathbb{P}(X \geq 0) = 1$, alors on a :*
$$\mathbb{E}[X] \geq 0.$$
3. ***Croissance***. *Soit X et Y deux v.a. réelles intégrables telles que $X \geq Y$ p.s., c'est-à-dire telles que $\mathbb{P}(X \geq Y) = 1$, alors on a :*
$$\mathbb{E}[X] \geq \mathbb{E}[Y].$$

On donne maintenant trois inégalités (voir aussi l'exercice II.18 pour une variante de l'inégalité de Tchebychev).

Proposition II.24. *Soit X une v.a. réelle.*

- ***Inégalité de Tchebychev*** *: Soit $a > 0$. On a :*
$$\boxed{\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}}.$$
- ***Inégalité de Jensen*** *: On suppose que X est intégrable. Soit φ une fonction* ***convexe****. Si $\mathbb{E}[\varphi(X)]$ existe alors on a :*
$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$
- ***Inégalité de de Cauchy-Schwarz*** *: Soit (X, Y) un couple de v.a. réelles. On suppose que X^2 et Y^2 sont intégrables. Alors XY est intégrable et on a :*
$$\boxed{\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].}$$

Dans une première lecture, on pourra omettre la démonstration de cette proposition qui est reportée à la fin du paragraphe.

Remarque II.25. La fonction $\varphi(x) = x^2$ est convexe. Soit X une v.a. réelle intégrable. On déduit de l'inégalité de Jensen que :

$$\boxed{\mathbb{E}[X]^2 \leq \mathbb{E}[X^2].}$$

$\Diamond$

On déduit de l'inégalité de Tchebychev le corollaire suivant.

Corollaire II.26. *Soit X une v.a. positive p.s. ($\mathbb{P}(X \geq 0) = 1$). Si $\mathbb{E}[X] = 0$ alors X est nulle p.s. ($\mathbb{P}(X = 0) = 1$).*

Démonstration. On déduit de l'inégalité de Tchebychev que pour tout entier $n \geq 1$, on a $\mathbb{P}(X \geq 1/n) = 0$. Par convergence monotone (cf. proposition I.2), on a $\mathbb{P}(X > 0) = 0$. Comme $\mathbb{P}(X \geq 0) = 1$, cela implique que p.s. $X = 0$. □

Démonstration de la proposition II.24. On a vu dans l'exemple II.22 que $\mathbb{P}(|X| \geq a) = \mathbb{E}\left[\mathbf{1}_{\{|X|\geq a\}}\right]$. Soit $\omega \in \Omega$.

- Si $|X(\omega)| \geq a$, alors on a $\mathbf{1}_{\{|X|\geq a\}}(\omega) = 1 \leq \dfrac{X(\omega)^2}{a^2}$.
- Si $|X(\omega)| < a$, alors on a $\mathbf{1}_{\{|X|\geq a\}}(\omega) = 0 \leq \dfrac{X(\omega)^2}{a^2}$.

Dans tous les cas, on a $\mathbf{1}_{\{|X|\geq a\}}(\omega) \leq \dfrac{X(\omega)^2}{a^2}$. Par croissance de l'espérance, on obtient l'inégalité de Tchebychev.

Pour la démonstration de l'inégalité de Jensen, on renvoie à l'exercice II.19.

Pour la démonstration de l'inégalité de Cauchy-Schwarz, on remarque que si $x, y \in \mathbb{R}$, alors $|xy| \leq (x^2 + y^2)/2$. Donc on a $|XY| \leq (X^2 + Y^2)/2$. Comme X^2 et Y^2 sont intégrables, par linéarité, $(X^2 + Y^2)/2$ est intégrable. On en déduit que $|XY|$ et XY sont intégrables. Si $\mathbb{E}[Y^2] = 0$, cela implique que $Y = 0$ p.s. d'après le corollaire II.26 (qui découle de 1) et donc on a p.s. $XY = 0$. L'inégalité de la proposition est alors triviale. On suppose que $\mathbb{E}[Y^2] > 0$. Par positivité et linéarité de l'espérance, on a, pour $\lambda \in \mathbb{R}$:

$$0 \leq \mathbb{E}[(X - \lambda Y)^2] = \mathbb{E}[X^2] - 2\lambda\mathbb{E}[XY] + \lambda^2\mathbb{E}[Y^2].$$

Le membre de droite est minimal pour $\lambda = \mathbb{E}[XY]/\mathbb{E}[Y^2]$. On obtient alors :

$$0 \leq \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]},$$

ce qui donne l'inégalité recherchée. □

II.8 Espérance d'une variable aléatoire discrète

La formule suivante du calcul de l'espérance d'une fonction d'une variable aléatoire discrète est très utile en pratique.

Proposition II.27.

1. *Soit φ une fonction réelle mesurable, soit X une v.a.d. réelle. On note Δ le support de sa loi. Si $\sum_{x\in\Delta} |\varphi(x)|\,\mathbb{P}(X=x) < \infty$ (on dit que $\varphi(X)$ est intégrable), on a :*

$$\boxed{\mathbb{E}[\varphi(X)] = \sum_{x\in\Delta} \varphi(x)\mathbb{P}(X=x).} \tag{II.5}$$

2. *La formule ci-dessus reste vraie si X est une v.a.d. à valeurs dans $\mathbb{R}^k$ et si φ est une fonction mesurable de $\mathbb{R}^k$ dans $\mathbb{R}^l$ avec la convention que $|\cdot|$ représente la norme euclidienne.*

Dans la proposition, on écrit parfois les sommations sur $\mathbb{R}$ ou $\mathbb{R}^d$ au lieu des sommations sur Δ.

Remarque. L'espérance et l'intégration usuelle partagent de nombreuses propriétés : linéarité, positivité, croissance. En effet l'espérance d'une variable aléatoire X ou d'une fonction de cette variable aléatoire $\varphi(X)$ est l'intégrale sur $\mathbb{R}$ par rapport, non pas à la mesure de Lebesgue, mais par rapport à la probabilité image $\mathrm{P_X}$, de la fonction x ou $\varphi(x)$. En particulier seule intervient la probabilité image. Ainsi si X et Y ont même loi et si $\varphi(X)$ est intégrable, alors $\varphi(Y)$ est intégrable et on a $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$. ◊

Démonstration. On démontre dans un premier temps la propriété 2. On pose $Z = \varphi(X)$. Il suffit donc de vérifier que :

$$\sum_{z\in\mathbb{R}^l} |z|\,\mathbb{P}(Z=z) = \sum_{x\in\mathbb{R}^k} |\varphi(x)|\,\mathbb{P}(X=x),$$

et si ces sommes sont finies, alors :

$$\sum_{z\in\mathbb{R}^l} z\mathbb{P}(Z=z) = \sum_{x\in\mathbb{R}^k} \varphi(x)\mathbb{P}(X=x).$$

La loi de la v.a.d. Z est donnée par la proposition II.17. Donc on a :

$$\sum_{z\in\mathbb{R}^l} |z|\,\mathbb{P}(Z=z) = \sum_{z\in\mathbb{R}^l} |z| \sum_{x\in\mathbb{R}^k} \mathbf{1}_{\{\varphi(x)=z\}}\mathbb{P}(X=x).$$

Les sommes comportent un nombre au plus dénombrable de termes non nuls, et tous les termes sont positifs. On peut donc intervertir les deux signes sommes, et il vient

$$\begin{aligned}\sum_{z\in\mathbb{R}^l} |z|\,\mathbb{P}(Z=z) &= \sum_{x\in\mathbb{R}^k}\sum_{z\in\mathbb{R}^l} |z|\,\mathbf{1}_{\{\varphi(x)=z\}}\mathbb{P}(X=x)\\ &= \sum_{x\in\mathbb{R}^k} |\varphi(x)|\,\mathbb{P}(X=x).\end{aligned}$$

Enfin, si cette dernière quantité est finie, on peut reproduire la même démonstration en enlevant les normes $|\cdot|$. Cette fois l'interversion des sommes est justifiée car les sommes sont absolument convergentes.

La propriété 1 est un cas particulier de 2. □

À titre d'exemple on donne la démonstration de la proposition II.23 pour le cas particulier des v.a. discrètes.

Démonstration de la proposition II.23. Pour montrer la propriété 1 de linéarité, on pose $Z=(X,Y)$ et pour $z=(x,y)$, $\varphi(z)=\alpha x+\beta y$. On note Δ le support de la loi de Z. On déduit de la proposition II.27 que :

$$\begin{aligned}\mathbb{E}[|\alpha X+\beta Y|] &= \sum_{(x,y)\in\Delta} |\alpha x+\beta y|\,\mathbb{P}(X=x,Y=y)\\ &\le \sum_{(x,y)\in\Delta} (|\alpha|\,|x|+|\beta|\,|y|)\mathbb{P}(X=x,Y=y)\\ &\le |\alpha| \sum_{(x,y)\in\Delta} |x|\,\mathbb{P}(X=x,Y=y) + |\beta| \sum_{(x,y)\in\Delta} |y|\,\mathbb{P}(X=x,Y=y).\end{aligned}$$

Il découle de la formule des lois marginales que :

$$\mathbb{E}[|\alpha X+\beta Y|] \le |\alpha|\,\mathbb{E}[|X|] + |\beta|\,\mathbb{E}[|Y|].$$

Cette dernière quantité est finie car X et Y sont intégrables. On en déduit que $\alpha X+\beta Y$ est intégrable. Enfin, en utilisant une deuxième fois la proposition II.27 et la formule des lois marginales, on obtient :

$$\begin{aligned}\mathbb{E}[\alpha X+\beta Y] &= \sum_{(x,y)\in\Delta} (\alpha x+\beta y)\mathbb{P}(X=x,Y=y)\\ &= \alpha \sum_{(x,y)\in\Delta} x\mathbb{P}(X=x,Y=y) + \beta \sum_{(x,y)\in\Delta} y\mathbb{P}(X=x,Y=y)\\ &= \alpha\mathbb{E}[X]+\beta\mathbb{E}[Y].\end{aligned}$$

La propriété 2 découle de (II.3) et du fait que pour une v.a. positive $\mathbb{P}(X = x) = 0$ si $x < 0$.

La propriété 3 découle de la propriété 2 avec $Z = X - Y$ et de la linéarité de l'espérance. □

On a également des formules de décomposition.

Exercice II.5.
Soit X et Y des v.a.d. On suppose que Y est à valeurs dans $\mathbb{R}^d$, et on note Δ_Y le support de sa loi. Soit φ une fonction mesurable telle que $\varphi(X)$ soit intégrable. Montrer que :

$$\boxed{\mathbb{E}[\varphi(X)] = \sum_{y \in \Delta_Y} \mathbb{E}\left[\varphi(X), Y = y\right],} \tag{II.6}$$

où par convention $\mathbb{E}\left[\varphi(X), Y = y\right] = \mathbb{E}\left[\varphi(X)\mathbf{1}_{\{Y=y\}}\right]$.

△

Correction II.5. On suppose que X est à valeurs dans $\mathbb{R}^l$. En utilisant la formule de décomposition (II.1), il vient

$$\begin{aligned}
\mathbb{E}[\varphi(X)] &= \sum_{x \in \mathbb{R}^l} \varphi(x)\mathbb{P}(X = x) \\
&= \sum_{x \in \mathbb{R}^l} \varphi(x) \sum_{y \in \Delta_Y} \mathbb{P}(X = x, Y = y) \\
&= \sum_{y \in \Delta_Y} \left(\sum_{z \in \Delta_Y, x \in \mathbb{R}^l} \varphi(x)\mathbf{1}_{\{z=y\}}\mathbb{P}(X = x, Y = z) \right) \\
&= \sum_{y \in \Delta_Y} \mathbb{E}\left[\varphi(X)\mathbf{1}_{\{Y=y\}}\right].
\end{aligned}$$

▲

Exercice II.6.
Soit X et Y des v.a.d. indépendantes. On note Δ_Y le support de la loi de Y. Soit φ une fonction mesurable bornée. Montrer que :

$$\boxed{\mathbb{E}[\varphi(X, Y)] = \sum_{y \in \Delta_Y} \mathbb{E}\left[\varphi(X, y)\right] \mathbb{P}(Y = y).}$$

△

II.9 Variance et Covariance

L'écart type et son carré, la variance, représentent l'écart ou la variation par rapport à la moyenne. Ils permettent de quantifier la dispersion des valeurs possibles par rapport à la moyenne. Ces notions sont très utilisées en calcul des probabilités mais aussi en statistique.

On dit qu'une v.a. réelle X est de **carré intégrable** si $\mathbb{E}[X^2]$ est fini. Il découle de la remarque II.25 ou de la majoration $2|x| \leq 1+x^2$, que toute variable de carré intégrable est intégrable.

Définition II.28. *Soit X une v.a. réelle de carré intégrable. On définit la **variance** de X, $\mathrm{Var}(X)$, par :*

$$\boxed{\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].}$$

*L'**écart type** est défini par $\sqrt{\mathrm{Var}(X)}$.*

La variance est **positive**. Le terme $\mathbb{E}[(X - \mathbb{E}[X])^2]$ est bien défini car les quantités $\mathbb{E}[X^2]$ et $\mathbb{E}[X]$ sont finies. Par linéarité, on a :

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Comme la variance est positive, on obtient que $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$. On retrouve ainsi le résultat de la remarque II.25.

Proposition II.29. *Soit X une v.a. de carré intégrable.*

1. *Soit $a, b \in \mathbb{R}$. On a :*
$$\boxed{\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).}$$
2. *Si $\mathrm{Var}(X) = 0$, alors il existe $a \in \mathbb{R}$ tel que p.s. $X = a$. Autrement dit une v.a. de variance nulle est p.s. constante.*

Démonstration. La propriété 1 est une conséquence directe de la linéarité de l'espérance. Pour démontrer la propriété 2, on remarque que la v.a. $(X - \mathbb{E}[X])^2$ est positive et d'espérance nulle. On déduit du corollaire II.26 que $\mathrm{Var}(X) = 0$ implique que p.s. $X = \mathbb{E}[X]$. □

Définition II.30. *Soit (X, Y) un couple de v.a. réelles de carré intégrable (et donc XY est intégrable grâce à l'inégalité de Cauchy-Schwarz). On définit la* ***covariance*** *de X et Y, $\mathrm{Cov}(X, Y)$, par :*

$$\boxed{\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].}$$

En particulier on a $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. La variance définit une forme quadratique sur l'espace vectoriel des v.a. de carré intégrable. La covariance est la forme bilinéaire associée à cette forme quadratique. On a, par linéarité de l'espérance :

$$\boxed{\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).} \tag{II.7}$$

II.10 Indépendance (II)

On donne la définition de l'indépendance pour des v.a. quelconques.

Définition II.31. *Deux variables aléatoires X et Y (réelles ou vectorielles) sont* ***indépendantes*** *si pour* ***toutes*** *fonctions réelles mesurables* ***bornées*** *f et g, on a :*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \tag{II.8}$$

On déduit de (II.8), avec $f = \mathbf{1}_A$ et $g = \mathbf{1}_B$, que si X et Y sont indépendants alors pour tous boréliens A et B, on a $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ et donc les évènements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants. La réciproque est vraie en général. Pour les v.a.d., c'est une conséquence du lemme II.33.

La définition d'indépendance s'étend à une famille quelconque de v.a.

Définition II.32.

1. *Soit $X_1, \ldots, X_n$, n variables aléatoires (réelles ou vectorielles). Elles sont* ***indépendantes*** *si pour* ***toutes*** *fonctions réelles mesurables* ***bornées*** *$f_1, \ldots, f_n$, on a :*

$$\boxed{\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)].} \tag{II.9}$$

2. *Soit $(X_i, i \in I)$ une famille quelconque de variables aléatoires réelles ou vectorielles. Elles sont indépendantes si pour toute famille* ***finie*** *d'indices $J \subset I$, les variables aléatoires $(X_i, i \in J)$ sont indépendantes.*

Lemme II.33. *Les deux définitions de l'indépendance II.11 (resp. II.12) et II.32-1 (resp. II.32-2) sont équivalentes pour les v.a. discrètes.*

Démonstration. Soit $X_1, \dots, X_n$ des v.a. discrètes indépendantes au sens de la définition II.32-1. En prenant $f_i(x) = \mathbf{1}_{\{x=x_i\}}$, on déduit de la définition II.32-1 et de l'exemple II.22 que : pour tout $(x_1, \dots, x_n)$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)] = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

Les v.a.d. sont donc indépendantes au sens de la définition II.11.

Soit $X_1, \dots, X_n$ des v.a. discrètes indépendantes au sens de la définition II.11. On note Δ_i, le support de la loi de X_i. Soit $f_1, \dots, f_n$ des fonctions mesurables bornées. On déduit de la formule de l'espérance pour les v.a.d. vectorielles, que :

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \sum_{x_1 \in \Delta_1, \dots, x_n \in \Delta_n} \left(\prod_{i=1}^n f_i(x_i)\right) \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Comme les v.a.d. sont indépendantes au sens de la définition II.11, on a :

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] &= \sum_{x_1 \in \Delta_1, \dots, x_n \in \Delta_n} \prod_{i=1}^n f_i(x_i)\, \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^n \sum_{x_i \in \Delta_i} f_i(x_i)\, \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^n \mathbb{E}[f_i(X_i)].
\end{aligned}$$

Les v.a.d. sont donc indépendantes au sens de la définition II.32-1.

On en déduit alors que les définitions II.12 et II.32-2 sont équivalentes. □

Remarque II.34. Si X et Y sont des v.a.d. de carré intégrable et indépendantes, alors on déduit de la démonstration du lemme II.33 que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. On admet que le résultat reste vrai pour des v.a. réelles indépendantes quelconques de carré intégrable. Donc, si X et Y sont indépendants et de carré intégrable, on a $\mathrm{Cov}(X, Y) = 0$. **La réciproque est fausse** en général. Voir le contre-exemple de l'exercice II.7. ◊

Exercice II.7.
Soit X une variable aléatoire de loi uniforme sur $\{1,\dots,6\}$. On pose $Y = \mathbf{1}_{\{X\in\{1,6\}\}}$. Montrer que $\mathrm{Cov}(X,Y) = 0$, mais que X et Y ne sont pas indépendants. △

Proposition II.35. *Soit X et Y deux variables aléatoires réelles indépendantes. On les suppose de carré intégrable. On a :*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Démonstration. Cela découle de la formule (II.7) et de la remarque II.34. □

Le corollaire suivant est immédiat.

Corollaire II.36. *Soit $X_1,\dots,X_n$, n v.a.* ***indépendantes et de carré intégrable****. On a :*

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).}$$

On donne une application de cette égalité dans le paragraphe suivant.

Exercice II.8.
Soit X et Y des v.a. indépendantes et soit g et f des fonctions mesurables réelles. Montrer que les v.a. $g(X)$ et $f(Y)$ sont indépendantes. △

Exemple. On démontre une variante de la formule de décomposition (voir aussi (II.6)). Soit X, Y deux v.a. indépendantes réelles ou vectorielles. On suppose de plus que Y est une v.a. discrète et on note Δ le support de la loi de Y. Soit φ une fonction mesurable bornée. On a la formule de décomposition suivante :

$$\boxed{\mathbb{E}[\varphi(X,Y)] = \sum_{y\in\Delta} \mathbb{E}[\varphi(X,y)]\mathbb{P}(Y=y).} \tag{II.10}$$

En effet, on a $\sum_{y\in\Delta} \mathbf{1}_{\{Y=y\}} = \mathbf{1}$ p.s. Cela implique :

$$\mathbb{E}[\varphi(X,Y)] = \mathbb{E}\left[\sum_{y\in\Delta} \varphi(X,y)\mathbf{1}_{\{Y=y\}}\right].$$

On peut intervertir le signe somme et l'espérance. Ce résultat découle du théorème de convergence dominée V.3 que l'on verra au chapitre V, voir l'exemple V.4. En utilisant l'indépendance, il vient :

$$\mathbb{E}[\varphi(X,Y)] = \sum_{y\in\Delta} \mathbb{E}\left[\varphi(X,y)\mathbf{1}_{\{Y=y\}}\right] = \sum_{y\in\Delta} \mathbb{E}[\varphi(X,y)]\mathbb{P}(Y=y).$$

$\Diamond$

II.11 Loi faible des grands nombres

On lance N fois une pièce. Soit $N(P)$ le nombre d'apparitions du côté pile. On observe que la fréquence empirique d'apparitions du côté pile $\dfrac{N(P)}{N}$ "converge" vers la probabilité d'obtenir pile. Le résultat suivant justifie cette intuition.

On modélise les lancers par une suite $(X_n, n \in \mathbb{N}^*)$ de v.a.d. indépendantes de loi de Bernoulli de paramètre p. Si $X_n = 1$, cela signifie que l'on a obtenu un pile au n-ième lancer. $S_n = \sum_{i=1}^n X_i$ représente le nombre de fois où pile est apparu en n lancers, et S_n/n représente la fréquence empirique ou la moyenne empirique. La proposition suivante assure que la probabilité pour que $\left|\frac{S_n}{n} - p\right|$ soit plus grand que ε est proche de 0 pour n grand. Ce résultat est valable pour tout $\varepsilon > 0$. On dit que S_n/n **converge en probabilité** vers p.

Proposition II.37. Loi faible des grands nombres. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a.* ***indépendantes*** *de* ***carré intégrable*** *et de* ***même loi****. On note $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^n X_i$ la* ***moyenne empirique****. La moyenne empirique converge en probabilité vers $\mathbb{E}[X_1]$: pour tout $\varepsilon > 0$,*

$$\boxed{\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| > \varepsilon\right) = 0.}$$

On démontrera un résultat plus fort et plus général au paragraphe V.4 : la loi forte des grands nombres.

Démonstration. On applique l'inégalité de Tchebychev et il vient en utilisant l'indépendance des v.a. X_i (proposition II.35) :

$$\begin{aligned}\mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| > \varepsilon\right) &\leq \mathbb{E}\left[\left(\bar{X}_n - \mathbb{E}[X_1]\right)^2\right]/\varepsilon^2 \\ &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i/n\right)/\varepsilon^2 \\ &= \mathrm{Var}(X_1)/n\varepsilon^2.\end{aligned}$$

Comme la variance de X_1 est finie, on en déduit le résultat. □

II.12 Fonctions génératrices

On introduit maintenant un outil pratique pour l'étude des variables aléatoires à valeurs dans l'ensemble des entiers naturels, $\mathbb{N}$. On donne un bref résumé des définitions et propriétés sur les séries entières au paragraphe II.15.

Définition II.38. *Soit X une v.a.d. à valeurs dans $\mathbb{N}$. On appelle fonction **génératrice** de X la série entière :*

$$\boxed{\phi_X(z) = \sum_{k=0}^{\infty} z^k \mathbb{P}(X = k) = \mathbb{E}\left[z^X\right],} \quad z \in [-1, 1].$$

La série est normalement convergente pour $z \in [-1, 1]$. En effet on a pour tout $z \in [-1, 1]$ et $k \in \mathbb{N}$, $\left|z^k \mathbb{P}(X = k)\right| \leq \mathbb{P}(X = k)$ et $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$. Le rayon de convergence de la série est donc supérieur ou égal à 1. Cela implique entre autre que ϕ_X est de classe C^∞ sur $]-1, 1[$ au moins, et les dérivées d'ordre n sont données par :

$$\boxed{\phi_X^{(n)}(z) = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} z^{k-n} \mathbb{P}(X = k);} \quad z \in]-1, 1[.$$

Remarque. La fonction génératrice ne dépend que de la loi de X. Ainsi deux v.a.d. à valeurs dans $\mathbb{N}$ ayant **même loi** ont **même fonction génératrice**. ◊

Proposition II.39. *La fonction génératrice de la v.a.d. X à valeurs dans $\mathbb{N}$ caractérise la loi de X.*

Démonstration. Comme X est une v.a.d. à valeurs dans $\mathbb{N}$, sa loi est caractérisée par la famille $(\mathbb{P}(X = n), n \in \mathbb{N})$. La valeur de la dérivée d'ordre n de ϕ_X en 0 est $\phi_X^{(n)}(0) = n!\mathbb{P}(X = n)$. On déduit donc à partir de ϕ_X la loi de X. □

On appelle $\mathbb{E}[X^n]$ le moment d'ordre n de la v.a.d. X. On remarque qu'il est toujours défini pour des v.a.d. positives, mais qu'il peut prendre la valeur $+\infty$. On peut facilement retrouver les moments de X à l'aide de la fonction génératrice. Les termes de la série définissant $\phi_X^{(n)}(z)$ sont tous positifs. On peut donc définir $\phi_X^{(n)}(1)$ comme la limite croissante de $\phi_X^{(n)}(z)$ quand $z \uparrow 1$. La valeur de cette limite peut être égale à $+\infty$.

Proposition II.40. *Pour tout entier $n \geq 1$, on a :*

$$\mathbb{E}\left[\prod_{l=0}^{n-1}(X - l)\right] = \phi_X^{(n)}(1).$$

En particulier $\mathbb{E}[X] = \phi'_X(1)$.

Démonstration. On a par convergence monotone que :

$$\phi_X^{(n)}(1) = \sum_{k \geq n} \prod_{l=0}^{n-1}(k - l)\mathbb{P}(X = k).$$

Cette dernière expression est exactement $\mathbb{E}\left[\prod_{l=0}^{n-1}(X - l)\right]$ (reconnaître $\mathbb{E}[X]$ si $n = 1$). □

Exercice II.9.
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ de carré intégrable. Montrer que $\phi'_X(1)$ et $\phi''_X(1)$ sont finis et que $\mathrm{Var}(X) = \phi''_X(1) + \phi'_X(1) - \phi'_X(1)^2$. △

Les fonctions génératrices permettent parfois de calculer facilement des lois.

Proposition II.41. *Soit $X_1, \ldots, X_n$ des v.a.d. définies à valeurs dans $\mathbb{N}$. On suppose qu'elles sont indépendantes, alors la fonction génératrice de la somme $X_1 + \cdots + X_n$ est le produit des fonctions génératrices :*

$$\phi_{X_1+\cdots+X_n}(z) = \prod_{i=1}^{n} \phi_{X_i}(z), \quad z \in [-1, 1].$$

Démonstration. Soit $z \in [-1,1]$. La fonction définie sur $\mathbb{R}$ par $f_z(x) = z^x \mathbf{1}_{\{x\in\mathbb{N}\}}$ est bornée et mesurable. En utilisant la définition II.32 sur l'indépendance des variables aléatoires, on a :

$$\begin{aligned}\phi_{X_1+\cdots+X_n}(z) &= \mathbb{E}\left[z^{X_1}\cdots z^{X_n}\right] \\ &= \mathbb{E}[f_z(X_1)\dots f_z(X_n)] \\ &= \prod_{i=1}^{n}\mathbb{E}[f_z(X_i)] \\ &= \prod_{i=1}^{n}\mathbb{E}\left[z^{X_i}\right] = \prod_{i=1}^{n}\phi_{X_i}(z).\end{aligned}$$

□

Exemple. On calcule la fonction génératrice $\phi_{\mathcal{B}(n,p)}$ de la loi binomiale $\mathcal{B}(n,p)$. Soit des v.a.d. $X_1,\dots,X_n$ indépendantes et de même loi de Bernoulli de paramètre p. La loi de $S_n = X_1 + \cdots + X_n$ est la loi binomiale $\mathcal{B}(n,p)$. Grâce à la proposition ci-dessus, on a : pour $z \in [-1,1]$,

$$\phi_{\mathcal{B}(n,p)}(z) = \phi_{S_n}(z) = \phi_{X_1}(z)^n = \mathbb{E}\left[z^{X_1}\right]^n = (1-p+pz)^n.$$

On peut alors retrouver la loi de S_n à l'aide de la formule du binôme. ◊

La proposition suivante assure qu'il est facile de calculer la loi de la somme d'un nombre aléatoire de variables aléatoires indépendantes.

Proposition II.42. *Soit $(X_n, n \in \mathbb{N})$ une suite de v.a.d. à valeurs dans $\mathbb{N}$, indépendantes et identiquement distribuées. Soit N une v.a.d. à valeurs entières, indépendante de la suite précédente. On considère $S = \sum_{n=1}^{N} X_n$ et $S = 0$ si $N = 0$. Alors on a :*

$$\phi_S(z) = \phi_N \circ \phi_{X_1}(z), \quad z \in [-1,1].$$

Démonstration. On utilise la formule de décomposition (II.6), et on reprend la démonstration de la proposition II.41. Soit $z \in [-1,1]$. On a :

$$\phi_S(z) = \mathbb{E}\left[z^S\right] = \sum_{n=0}^{\infty}\mathbb{E}\left[z^S, N=n\right].$$

On distingue le cas $N = 0$:

$$\mathbb{E}\left[z^S\right] = \mathbb{P}(N=0) + \sum_{n=1}^{\infty} \mathbb{E}\left[z^{X_1}\cdots z^{X_n}, N=n\right].$$

Par indépendance, il vient :

$$\begin{aligned}\mathbb{E}\left[z^S\right] &= \mathbb{P}(N=0) + \sum_{n=1}^{\infty}\prod_{i=1}^{n}\mathbb{E}\left[z^{X_i}\right]\mathbb{P}(N=n)\\ &= \sum_{n=0}^{\infty}\left[\phi_{X_1}(z)\right]^n\mathbb{P}(N=n)\\ &= \phi_N\circ\phi_{X_1}(z).\end{aligned}$$

On a utilisé le fait que $\phi_{X_1} \in [-1,1]$ si $z \in [-1,1]$ pour la dernière égalité. □

II.13 Indépendance (III)

On généralise la définition des fonctions génératrices aux v.a.d. vectorielles à valeurs dans $\mathbb{N}^d$. Ceci permet de donner une nouvelle caractérisation de l'indépendance pour des variables aléatoires à valeurs entières.

Définition II.43. *Soit $X = (X_1, \ldots, X_d)$ une v.a.d. à valeurs dans $\mathbb{N}^d$. On appelle fonction génératrice de X la série : pour $z = (z_1, \ldots, z_d) \in [-1,1]^d$,*

$$\boxed{\phi_X(z) = \sum_{k=(k_1,\ldots,k_d)\in\mathbb{N}^d} z_1^{k_1}\cdots z_d^{k_d}\mathbb{P}(X=k) = \mathbb{E}\left[z_1^{X_1}\cdots z_d^{X_d}\right].}$$

La série est normalement convergente pour $z \in [-1,1]^d$, et la fonction ϕ_X est de classe C^∞ sur $]-1,1[^d$ au moins.

Proposition II.44. *Soit $X = (X_1, \ldots, X_d)$ une v.a.d. vectorielle à valeurs dans $\mathbb{N}^d$.*

1. *La fonction génératrice de X caractérise la loi de X.*
2. *Si on connaît la fonction génératrice du vecteur X, on peut en déduire la fonction génératrice de la coordonnée X_i. On a :*

$$\phi_{X_i}(z_i) = \phi_X(z) \quad \text{où } z = (z_1,\ldots,z_d),\ z_i \in [-1,1] \text{ et } z_j = 1 \quad \forall j \neq i.$$

3. Les v.a.d. $X_1, \dots, X_d$ à valeurs entières sont indépendantes si et seulement si :

$$\boxed{\phi_{X_1,\dots,X_d}(z) = \prod_{i=1}^{d} \phi_{X_i}(z_i)} \quad \textit{pour tout} \quad z = (z_1, \dots, z_d) \in [-1,1]^d.$$

Démonstration. 1. En différenciant la fonction génératrice du vecteur X, il vient :

$$\frac{\partial^{n_1+\dots+n_d}\phi_X(z_1,\dots,z_d)}{\partial^{n_1} z_1 \cdots \partial^{n_d} z_d}(0) = n_1! \cdots n_d!\, \mathbb{P}(X_1 = n_1, \dots, X_d = n_d).$$

On en déduit que ϕ_X caractérise la loi de X.

2. Cela découle de la définition de ϕ_X et de ϕ_{X_i}.

3. Le sens direct est une conséquence de la définition II.32. Pour la réciproque on vérifie pour tout $n_1, \dots, n_d \geq 0$, l'égalité $\mathbb{P}(X_1 = n_1, \dots, X_d = n_d) = \prod_{i=1}^{d} \mathbb{P}(X_i = n_i)$ en différenciant n_i fois par rapport à z_i la relation $\phi_{X_1,\dots,X_d}(z) = \prod_{i=1}^{d} \phi_{X_i}(z_i)$ et en évaluant en $z = 0$. □

II.14 Lois conditionnelles et espérances conditionnelles

La notion de loi conditionnelle est une extension de la notion de probabilité conditionnelle.

Définition II.45. *Soit X, Y deux v.a.d. définies sur le même espace probabilisé. La **loi conditionnelle** de Y sachant X notée $\mathcal{L}(Y|X)$ est caractérisée par $\mathbb{P}(Y = y|X = x)$ pour x et y appartenant respectivement au support de la loi de X et de la loi de Y.*

Plus généralement pour une v.a. Y quelconque et une v.a. discrète X, on définit la loi conditionnelle de Y sachant X par la famille des probabilités images $(\mathrm{P}_{Y|X=x}, x \in \Delta)$, où Δ est le support de la loi de X et $\mathrm{P}_{Y|X=x}(B) = \mathbb{P}(Y \in B|X = x)$ pour tout ensemble mesurable B.

Exemple II.46. Soit $(X_i, i \in \{1, \dots, n\})$ une suite de v.a.d. indépendantes de loi de Bernoulli de paramètre p. On pose $S_n = \sum_{i=1}^{n} X_i$. On calcule la loi conditionnelle de X_1 sachant S_n. On a par indépendance :

$$\mathbb{P}(X_1 = 1|S_n = k) = \frac{\mathbb{P}(X_1 = 1, X_1 + \cdots + X_n = k)}{\mathbb{P}(S_n = k)}$$
$$= \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 + \cdots + X_n = k-1)}{\mathbb{P}(S_n = k)}.$$

Comme $X_2 + \cdots + X_n$ suit une loi binomiale $\mathcal{B}(n-1, p)$, il vient pour $k \geq 1$,

$$\mathbb{P}(X_1 = 1|S_n = k) = \frac{p\, C_{n-1}^{k-1}\, p^{k-1}(1-p)^{n-k}}{C_n^k\, p^k(1-p)^{n-k}} = \frac{k}{n}.$$

Cette égalité est aussi vraie pour $k = 0$. De manière plus condensée, on écrira que :

$$\mathbb{P}(X_1 = 1|S_n) = \frac{S_n}{n}.$$

On vérifie que $\mathbb{P}(X_1 = 0|S_n) = 1 - \frac{S_n}{n}$. On dira que conditionnellement à S_n, la loi de X_1 est une loi de Bernoulli de paramètre S_n/n. ◊

Exercice II.10.
On reprend les notations de l'exemple II.46. On suppose $n \geq 2$. Calculer la loi conditionnelle de (X_1, X_2) sachant S_n. Les variables X_1 et X_2 sont-elles indépendantes conditionnellement à S_n ? △

Remarque. Soit X une v.a.d. et Y une v.a. réelle ou vectorielle indépendante de X. Alors la loi conditionnelle de Y sachant X est simplement la loi de Y. En effet on a $\mathbb{P}(Y \in B|X = x) = \mathbb{P}(Y \in B)$ pour tout x appartenant au support de la loi de X et pour tout borélien B. ◊

Soit Y une v.a. réelle de carré intégrable. On considère l'application $h : a \mapsto \mathbb{E}[(Y-a)^2]$. Comme $h(a) = \mathbb{E}[Y^2] + a^2 - 2a\mathbb{E}[Y]$, on en déduit que h atteint son minimum en $a = \mathbb{E}[Y]$. Ainsi l'espérance de Y apparaît comme la meilleure approximation de Y par une constante au sens quadratique. L'espérance conditionnelle de Y sachant une v.a.d. X, définie ci-dessous, peut être également vue comme la meilleure approximation de Y par une fonction de X au sens quadratique (voir la proposition II.51 pour un énoncé plus précis).

Définition II.47. *Soit X une v.a. discrète et Y une v.a. intégrable quelconque. On définit l'**espérance conditionnelle** de Y sachant X, notée $\mathbb{E}[Y|X]$ par la v.a. $\psi(X)$ où :*

$$\psi(x) = \frac{\mathbb{E}[Y\mathbf{1}_{\{X=x\}}]}{\mathbb{P}(X = x)}.$$

si $\mathbb{P}(X = x) > 0$ et $\psi(x) = 0$ sinon. Enfin, on note également $\mathbb{E}[Y|X = x] = \psi(x)$.

Comme Y est intégrable, la v.a. $Y\mathbf{1}_{\{X=x\}}$ est intégrable. La fonction $\psi(x)$ est donc bien définie.

Exemple. Soit $(X_i, i \in \{1,\dots,n\})$ une suite de v.a.d. indépendantes de loi de Bernoulli de paramètre p. On pose $S_n = \sum_{i=1}^n X_i$. On déduit de l'exemple II.46 que $\sum_{x_1\in\{0,1\}} x_1 \mathbb{P}(X_1 = x_1 | S_n = x) = x/n$. L'espérance conditionnelle de X_1 sachant S_n est donc $\mathbb{E}[X_1|S_n] = S_n/n$. On remarque que la loi conditionnelle de X_1 sachant S_n est la loi de Bernoulli de paramètre $p' = S_n/n$. L'espérance d'une variable aléatoire de Bernoulli de paramètre p' est p'. L'espérance conditionnelle de X_1 sachant S_n est donc $p' = S_n/n$. ◊

Exercice II.11.
Soit A et B deux évènements tels que $\mathbb{P}(B) \in]0,1[$. Calculer la loi conditionnelle de $\mathbf{1}_A$ sachant $\mathbf{1}_B$. Calculer $\mathbb{E}[\mathbf{1}_A|\mathbf{1}_B]$. △

On donne quelques propriétés de l'espérance conditionnelle.

Proposition II.48. *Soit X une v.a. discrète et Y une v.a. quelconque.*

1. *Soit φ une fonction mesurable telle que $\varphi(X,Y)$ soit intégrable. La v.a. $\mathbb{E}[\varphi(X,Y)|X]$ est intégrable, et on a :*

$$\boxed{\mathbb{E}\big[\mathbb{E}[\varphi(X,Y)|X]\big] = \mathbb{E}[\varphi(X,Y)].}$$

2. *Soit f une fonction mesurable telle que $f(Y)$ soit intégrable, soit g est une fonction réelle mesurable bornée, alors p.s. :*

$$\boxed{\mathbb{E}[g(X)f(Y)|X] = g(X)\mathbb{E}[f(Y)|X].}$$

   En particulier on a $\mathbb{E}[g(X)f(Y)] = \mathbb{E}\big[g(X)\mathbb{E}[f(Y)|X]\big]$.

3. *Soit f une fonction mesurable telle que $f(Y)$ soit intégrable. Si X et Y sont indépendants, alors on a p.s. $\mathbb{E}[f(Y)|X] = \mathbb{E}[f(Y)]$.*
4. *Soit g une fonction mesurable telle que $g(X)$ soit intégrable, alors on a $\mathbb{E}[g(X)|X] = g(X)$ p.s.*

Démonstration. On note Δ_X le support de la loi de X.

1. On pose $\psi(x) = \mathbb{E}[\varphi(X,Y)|X = x]$, et on remarque que par définition on a :

$$\psi(x) = \frac{\mathbb{E}[\varphi(X,Y)\mathbf{1}_{\{X=x\}}]}{\mathbb{P}(X=x)} = \frac{\mathbb{E}[\varphi(x,Y)\mathbf{1}_{\{X=x\}}]}{\mathbb{P}(X=x)} = \mathbb{E}[\varphi(x,Y)|X=x]. \qquad \text{(II.11)}$$

Il vient :

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\varphi(X,Y)|X]] = \mathbb{E}[\psi(X)] &= \sum_{x\in\Delta_X} \psi(x)\mathbb{P}(X=x) \\ &= \sum_{x\in\Delta_X} \mathbb{E}[\varphi(x,Y)\mathbf{1}_{\{X=x\}}] = \mathbb{E}[\varphi(X,Y)],\end{aligned}$$

car on peut intervertir le signe somme et l'espérance pour la dernière égalité (ce résultat découle du théorème de convergence dominée V.3).

2. On déduit de (II.11) que :

$$\mathbb{E}[g(X)f(Y)|X=x] = \mathbb{E}[g(x)f(Y)|X=x] = g(x)\mathbb{E}[f(Y)|X=x].$$

On a donc $\mathbb{E}[g(X)f(Y)|X] = g(X)\mathbb{E}[f(Y)|X]$, et en prenant l'espérance, on déduit de la propriété 1 que $\mathbb{E}[g(X)f(Y)] = \mathbb{E}\big[g(X)\mathbb{E}[f(Y)|X]\big]$.

3. Comme X et Y sont indépendants, il vient :

$$\mathbb{E}[f(Y)|X=x] = \frac{\mathbb{E}[f(Y)\mathbf{1}_{\{X=x\}}]}{\mathbb{P}(X=x)} = \mathbb{E}[f(Y)].$$

4. Le résultat découle de la définition de l'espérance conditionnelle. □

Enfin, on signale que l'espérance conditionnelle partage les mêmes propriétés que l'intégrale et l'espérance : linéarité, positivité, croissance. La démonstration de la proposition suivante est laissée en exercice (voir la démonstration dans le cas des v.a.d. de la proposition II.23).

Proposition II.49. *Soit X une v.a.d.*

1. ***Linéarité**. Soit Y et Z deux v.a. intégrables, soit $\alpha, \beta \in \mathbb{R}$. Alors on a p.s. :*

$$\mathbb{E}[\alpha Y + \beta Z|X] = \alpha\mathbb{E}[Y|X] + \beta\mathbb{E}[Z|X].$$

2. ***Positivité**. Soit Y une v.a. réelle positive alors p.s. $\mathbb{E}[Y|X] \geq 0$.*
3. ***Croissance**. Soit Y et Z deux v.a. réelles intégrables telles que $Y \geq Z$ p.s., alors p.s. $\mathbb{E}[Y|X] \geq \mathbb{E}[Z|X]$.*

On donne un lemme du type inégalité de Jensen pour l'espérance conditionnelle.

Lemme II.50. *Soit X une v.a. discrète. Si Y est une v.a. réelle de carré intégrable, alors $\mathbb{E}[Y|X]$ est de carré intégrable et on a $\mathbb{E}[Y|X]^2 \leq \mathbb{E}[Y^2|X]$.*

Démonstration. Soit x tel que $\mathbb{P}(X = x) > 0$. On déduit de l'exercice I.7, que $Q(\cdot) = \mathbb{P}(\cdot|X = x)$ est une probabilité. En particulier $\psi(x) = Q[Y] = \mathbb{E}[Y|X = x]$. On déduit de la remarque II.25 que $Q[Y]^2 \leq Q[Y^2]$. Cela signifie que $\mathbb{E}[Y|X]^2 \leq \mathbb{E}[Y^2|X]$. Comme $\mathbb{E}[Y^2|X]$ est intégrable, on en déduit que $\mathbb{E}[Y|X]$ est de carré intégrable. □

Exercice II.12.
Soit N une v.a.d. à valeurs dans $\mathbb{N}$. Soit $(X_k, k \in \mathbb{N})$ une suite de v.a. réelles de carré intégrable, de même loi, indépendantes et indépendantes de N. On pose $S_n = 0$ si $n = 0$ et $S_n = \sum_{k=1}^{n} X_k$ si $n \in \mathbb{N}^*$. Calculer $\mathbb{E}[S_N]$ et $\mathrm{Var}(S_N)$. △

Correction II.12. On a pour $n \in \mathbb{N}$:

$$\mathbb{E}[S_N|N = n] = \mathbb{E}\Big[\sum_{k=1}^{n} X_k|N = n\Big] = n\mathbb{E}[X_1|N = n] = n\mathbb{E}[X_1].$$

On en déduit donc $\mathbb{E}[S_N|N] = N\mathbb{E}[X_1]$ et

$$\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X_1].$$

Cette dernière égalité s'appelle équation de Wald. Le calcul de la variance est similaire. On a pour $n \in \mathbb{N}$,

$$\mathbb{E}[S_N^2|N = n] = \mathbb{E}[S_n^2] = \mathrm{Var}(S_n) + \mathbb{E}[S_n]^2 = n\,\mathrm{Var}(X_1) + n^2\mathbb{E}[X_1]^2.$$

On en déduit donc que $\mathbb{E}[S_N^2] = \mathbb{E}[N]\,\mathrm{Var}(X_1) + \mathbb{E}[N^2]\mathbb{E}[X_1]^2$, et donc

$$\mathrm{Var}(S_N) = \mathbb{E}[N]\,\mathrm{Var}(X_1) + \mathbb{E}[X_1]^2\,\mathrm{Var}(N).$$

▲

Exercice II.13.
Soit φ une fonction réelle convexe. Soit Y une v.a. intégrable. On suppose que $\varphi(Y)$ est positive ou intégrable. Montrer que $\varphi(\mathbb{E}[Y|X]) \leq \mathbb{E}[\varphi(Y)|X]$. On pourra s'inspirer de l'exercice II.19. △

On peut omettre la fin de ce paragraphe dans une première lecture.

On considère $\mathcal{L}^2(\Omega)$ l'ensemble des variables aléatoires réelles de carré intégrables. Pour $Y \in \mathcal{L}^2(\Omega)$, on pose $\|Y\| = \sqrt{\mathbb{E}[Y^2]}$. Si $\|Y\| = 0$, alors on en déduit

que $Y = 0$ p.s. Ainsi $\|\cdot\|$ n'est pas une norme sur $\mathcal{L}^2(\Omega)$, mais c'est une norme sur l'espace vectoriel L^2 défini comme $\mathcal{L}^2(\Omega)$ quotienté par la relation d'équivalence de l'égalité presque sûre. On considère le produit scalaire associé $(Y, Z) = \mathbb{E}[YZ]$. L'espace L^2 muni de ce produit scalaire est un espace de Hilbert, voir page 131.

On définit l'espérance d'une v.a. réelle Y de carré intégrable sachant la variable aléatoire X comme la projection orthogonale de Y sur le sous-espace vectoriel des fonctions de X de carré intégrable :

$$\mathcal{H} = \{\text{v.a. réelle } \varphi(X) \text{ de carré intégrable, } \varphi \text{ mesurable}\}.$$

(En fait l'espace $\mathcal{H}$ est un sous-espace fermé de L^2.)

On vérifie que, dans le cas où X est une v.a.d. et Y une v.a. réelle, cette définition coïncide avec la définition II.47. Pour cela, il suffit de vérifier la proposition suivante.

Proposition II.51. *Soit X et Y deux variables aléatoires. On suppose X discrète et Y de carré intégrable. Alors on a $\mathbb{E}[Y|X] \in \mathcal{H}$, où $\mathbb{E}[Y|X]$ est donné par la définition II.47. De plus $\mathbb{E}[Y|X]$ minimise $\mathbb{E}[(Y-W)^2]$ pour $W \in \mathcal{H}$ et $Y-\mathbb{E}[Y|X]$ est orthogonal à $\mathcal{H}$ ($\mathbb{E}[Y|X]$ s'interprète donc comme la projection orthogonale de Y sur $\mathcal{H}$) : $\mathbb{E}[(Y - \mathbb{E}[Y|X])W'] = 0$ pour tout $W' \in \mathcal{H}$. Enfin, s'il existe une variable aléatoire $Z \in \mathcal{H}$ telle que pour toute fonction g, où $g(X)$ est de carré intégrable, on a $\mathbb{E}[Zg(X)] = \mathbb{E}[Yg(X)]$, alors p.s. on a $Z = \mathbb{E}[Y|X]$.*

Démonstration. Comme $\mathbb{E}[Y|X] = \psi(X)$ et que $\mathbb{E}[Y|X]$ est de carré intégrable (voir lemme II.50), on a donc $\mathbb{E}[Y|X] \in \mathcal{H}$. Soit $W' \in \mathcal{H}$, on déduit du 3 de la proposition II.48 que :

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])W'] = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y|X])|X]W'] = 0.$$

Donc $Y - \mathbb{E}[Y|X]$ est orthogonal à $\mathcal{H}$.

En faisant intervenir le terme $\mathbb{E}[Y|X]$, il vient, pour $W \in \mathcal{H}$:

$$\begin{aligned}\mathbb{E}[(Y-W)^2] &= \mathbb{E}[((Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - W))^2] \\ &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - W)^2] \\ &\qquad + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - W)] \\ &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - W)^2],\end{aligned}$$

car $\mathbb{E}[Y|X] - W \in \mathcal{H}$. Pour minimiser $\mathbb{E}[(Y - W)^2]$, il faut donc minimiser $\mathbb{E}[(\mathbb{E}[Y|X] - W)^2]$. Le minimum est atteint pour $W = \mathbb{E}[Y|X]$.

Soit $Z = \varphi(X)$ une v.a.d. de carré intégrable telle que pour toute fonction g mesurable, où $g(X)$ est de carré intégrable, $\mathbb{E}[Zg(X)] = \mathbb{E}[Yg(X)]$. On a donc $\mathbb{E}[(\varphi(X)-\psi(X))g(X)] = 0$. On choisit $g(x) = \varphi(x)-\psi(x)$. On a bien $g(X)$ de carré intégrable. Il vient $\mathbb{E}[(\varphi(X)-\psi(X))^2] = 0$. On en déduit que p.s. $\varphi(X)-\psi(X) = 0$ c'est-à-dire p.s. $Z = \mathbb{E}[Y|X]$. □

II.15 Rappels sur les séries et les séries entières

Les définitions et propositions qui suivent sont valables sur $\mathbb{R}$ mais aussi sur tout espace de Banach.

Une série $\sum_{n\in\mathbb{N}} a_n$ est convergente si la suite $(b_n, n \in \mathbb{N})$ définie par $b_n = \sum_{k=0}^{n} a_k$ est convergente. La valeur de la série est la limite de la suite $(b_n, n \in \mathbb{N})$. On dit qu'une série $\sum_{n\in\mathbb{N}} a_n$ est **absolument** convergente si la série $\sum_{n\in\mathbb{N}} |a_n|$ est convergente.

Proposition II.52. *Toute série absolument convergente est convergente. En outre, si $\sum_{n\in\mathbb{N}} a_n$ est absolument convergente alors pour toute permutation σ de $\mathbb{N}$, la série $\sum_{n\in\mathbb{N}} a_{\sigma(n)}$ est convergente, et sa valeur est indépendante de σ.*

Si la série $\sum_{n\in\mathbb{N}} a_n$ à termes réels est convergente mais pas absolument convergente, alors pour toute valeur $x \in \overline{\mathbb{R}}$, on peut trouver une permutation σ telle que la série $\sum_{n\in\mathbb{N}} a_{\sigma(n)}$ converge vers x.

Proposition II.53 (Fubini)**.** *Si la série $\sum_{n\geq 0, p\geq 0} |a_{n,p}|$ est convergente, alors on a :*

$$\sum_{n\geq 0}\left(\sum_{p\geq 0} a_{n,p}\right) = \sum_{p\geq 0}\left(\sum_{n\geq 0} a_{n,p}\right).$$

Proposition II.54 (Convergence dominée)**.** *Soit $(a_{n,p}, n \geq 0, p \geq 0)$ une suite telle que pour tout $n \geq 0, p \geq 0$, $|a_{n,p}| \leq b_n$. On suppose de plus que pour tout $n \geq 0$, $\lim_{p\to\infty} a_{n,p}$ existe. Si la série à termes positifs $\sum_{n\geq 0} b_n$ est (absolument) convergente, alors on a :*

$$\lim_{p\to\infty} \sum_{n\geq 0} a_{n,p} = \sum_{n\geq 0} \lim_{p\to\infty} a_{n,p}.$$

On se place dorénavant sur $\mathbb{R}$ ou $\mathbb{C}$.

Soit $(a_n, n \in \mathbb{N})$ une suite de nombres complexes. La série entière associée à la suite $(a_n, n \in \mathbb{N})$ est la série $\sum_{n\in\mathbb{N}} a_n z^n$ où $z \in \mathbb{C}$. Le **rayon de convergence** de la série entière est défini par :

$$R = \sup\{r > 0; \sum_{n\in\mathbb{N}} |a_n|\, r^n \text{ est convergente}\},$$

avec la convention $\sup \emptyset = 0$. On rappelle qu'une série de fonctions $\sum_{n\geq 0} f_n(x)$ est **normalement** convergente sur un ensemble A, s'il existe une suite $(b_n, n \in \mathbb{N})$ telle que $|f_n(x)| \leq b_n$ pour tout $x \in A$, et la série $\sum_{n\in\mathbb{N}} b_n$ est (absolument) convergente.

Proposition II.55.

1. *Si $|z| > R$, alors la série $\sum_{n\in\mathbb{N}} a_n z^n$ diverge trivialement (la suite $(a_n z^n)$ n'est pas bornée).*
2. *Si $|z| < R$, alors la série $\sum_{n\in\mathbb{N}} a_n z^n$ est absolument convergente.*
3. *Soit $0 < r < R$. La série $\sum_{n\in\mathbb{N}} a_n z^n$ est normalement convergente sur $\{z \in \mathbb{C}; |z| \leq r\}$.*

Exercice II.14.
Si $\left|\dfrac{a_{n+1}}{a_n}\right|$ converge vers une limite ℓ, alors le rayon de convergence de la série entière $\sum_{n\in\mathbb{N}} a_n z^n$ est $1/\ell$. △

La formule d'Hadamard assure que le rayon de convergence R de la série entière $\sum_{n\geq 0} a_n z^n$ est $R = 1/\limsup_{n\to\infty} |a_n|^{1/n}$.

On donne quelques propriétés des séries entières.

Proposition II.56. *Soit $\sum_{n\in\mathbb{N}} a_n z^n$ et $\sum_{n\in\mathbb{N}} b_n z^n$ deux séries entières de rayons de convergence respectifs $R_a > 0$ et $R_b > 0$.*

***Somme** : La série entière $\sum_{n\in\mathbb{N}}(a_n + b_n)z^n$ est bien définie pour $|z| < R_a \wedge R_b$. De plus pour $|z| < R_a \wedge R_b$, on a :*

$$\sum_{n\in\mathbb{N}}(a_n + b_n)z^n = \left(\sum_{n\in\mathbb{N}} a_n z^n\right) + \left(\sum_{n\in\mathbb{N}} b_n z^n\right).$$

***Produit** : La série entière $\sum_{n\in\mathbb{N}}(\sum_{p+q=n} a_p b_q)z^n$ est bien définie pour $|z| < R_a \wedge R_b$. De plus pour $|z| < R_a \wedge R_b$, on a :*

$$\sum_{n\in\mathbb{N}}\Big(\sum_{p+q=n} a_p b_q\Big)z^n = \left(\sum_{n\in\mathbb{N}} a_n z^n\right)\left(\sum_{n\in\mathbb{N}} b_n z^n\right).$$

Dérivation *: Pour $|z| < R_a$, on note $S(z) = \sum_{n\in\mathbb{N}} a_n z^n$. La fonction S est dérivable à tout ordre sur $\{z; |z| < R_a\}$, et on a :*

$$S^{(p)}(z) = \sum_{n\geq p} \frac{n!}{(n-p)!} a_n z^{n-p}.$$

De plus le rayon de convergence de cette série est R_a.

Intégration *: La fonction S est intégrable sur $]-R_a, R_a[$ et on a : pour $z \in]-R_a, R_a[$,*

$$\int_0^z S(y)\, dy = \sum_{n\in\mathbb{N}} \frac{a_n}{n+1} z^{n+1}.$$

De plus le rayon de convergence de cette série est R_a.

II.16 Résumé

Soit X une variable aléatoire discrète (v.a.d.) réelle ou vectorielle. On appelle $\Delta = \{x; \mathbb{P}(X = x) > 0\}$ le **support** de sa loi. L'ensemble Δ est au plus dénombrable et on a $\mathbb{P}(X \in \Delta) = 1$.

- Soit φ une fonction mesurable. $\varphi(X)$ est une v.a.d. On dit que $\varphi(X)$ est **intégrable** si $\sum_{x\in\Delta} |\varphi(x)|\, \mathbb{P}(X = x) < +\infty$. Si $\varphi(X)$ est intégrable ou positive p.s., alors l'espérance de $\varphi(X)$ est définie par :

$$\mathbb{E}[\varphi(X)] = \sum_{x\in\Delta} \varphi(x)\mathbb{P}(X = x).$$

- Si A est un évènement, alors on a $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$.
- Formule de décomposition I : Soit B un évènement. On a la relation $\mathbb{P}(B) = \sum_{x\in\Delta} \mathbb{P}(B, X = x)$.
- Formule de décomposition II : Soit Y une v.a.d. et φ une fonction mesurable, telle que $\varphi(Y)$ soit intégrable. On a $\mathbb{E}[\varphi(Y)] = \sum_{x\in\Delta} \mathbb{E}\left[\varphi(Y)\mathbf{1}_{\{X=x\}}\right]$.

L'espérance est définie pour des variables quelconques intégrables.

- On a les propriétés suivantes de l'espérance : **linéarité, positivité, croissance**.
- Soit X, Y des v.a. réelles. On a les inégalités suivantes :
  - **Tchebychev** : $\mathbb{P}(X \geq a) \leq \mathbb{E}[X^2]/a^2$.
  - **Jensen** : $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ où φ est convexe, X intégrable et $\mathbb{E}[\varphi(X)]$ a un sens.
  - **Cauchy-Schwarz** : $\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$, où X et Y sont de carré intégrable.
- La **variance** d'une v.a. de carré intégrable est :

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2,$$

  et son **écart-type** $\sqrt{\mathrm{Var}(X)}$.
- La variance est une forme quadratique. Elle est toujours positive. On a $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.
- La forme bilinéaire associée est la **covariance** : $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ et $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.
- Les v.a. X et Y sont **indépendantes** si, pour toutes fonctions f, g mesurables bornées, on a $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$. Cette définition possède une extension pour une famille quelconque de variables aléatoires.
- Si X et Y sont indépendantes et de carré intégrable, alors on a $\mathrm{Cov}(X, Y) = 0$ et $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. **La réciproque est fausse**.

Autres propriétés des variables aléatoires discrètes.

- Deux v.a.d. X et Y sont **indépendantes** si et seulement si $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour tout (x, y) appartenant au support de la loi du couple (X, Y). Cette définition possède une extension pour une famille quelconque de variables aléatoires discrètes.
- Si X est une v.a.d. à valeurs entières, alors sa **fonction génératrice** est définie par :

$$\phi_X(z) = \mathbb{E}\left[z^X\right], \quad z \in [-1, 1].$$

- Deux v.a.d. X et Y à valeurs entières sont indépendantes si et seulement si :

$$\phi_{(X,Y)}(z_1, z_2) = \phi_X(z_1)\phi_Y(z_2) \quad \text{pour tout } (z_1, z_2) \in [-1, 1]^2.$$

- Si X et Y sont deux v.a.d. indépendantes, alors on a $\phi_{X+Y}(z) = \phi_X(z)\phi_Y(z)$.
- **Formule des lois marginales** : On peut calculer la loi de X et la loi de Y à partir de la loi du couple (X, Y) : $\mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y)$. **La réciproque est fausse**.
- Si X et Y sont deux v.a.d. indépendantes, alors la loi du couple (X, Y) est la **loi produit** : $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour tout couple (x, y).

– **Changement de variable** : Soit $Y = \varphi(X)$, où φ est mesurable. Y est une v.a.d. de loi $\mathbb{P}(Y = y) = \sum_{x\in\Delta} \mathbf{1}_{\{\varphi(x)=y\}}\mathbb{P}(X = x)$.

Loi conditionnelle et espérance conditionnelle sachant une v.a.d.

– Soit Y une v.a., et f une fonction mesurable réelle. Si $f(Y)$ est intégrable, l'**espérance conditionnelle** de $f(Y)$ sachant X est $\psi(X)$ où $\psi(x) = \mathbb{E}[f(Y)\mathbf{1}_{\{X=x\}}]/\mathbb{P}(X = x)$, et on a $\mathbb{E}[\mathbb{E}[f(Y)|X]] = \mathbb{E}[f(Y)]$.
– Si g est mesurable bornée, on a $\mathbb{E}[g(X)f(Y)|X] = g(X)\mathbb{E}[f(Y)|X]$ et bien sûr $\mathbb{E}[g(X)f(Y)] = \mathbb{E}[g(X)\mathbb{E}[f(Y)|X]]$.
– Si Y est indépendant de X, la loi conditionnelle de Y sachant X est la loi de Y : si $f(Y)$ est intégrable, alors on a $\mathbb{E}[f(Y)|X] = \mathbb{E}[f(Y)]$.
– L'espérance conditionnelle comme l'espérance possède les propriétés de linéarité, positivité et croissance.
– Si Y est une v.a. réelle de carré intégrable, alors $\mathbb{E}[Y|X]$ est de carré intégrable et on a $\mathbb{E}[Y|X]^2 \leq \mathbb{E}[Y^2|X]$.
– Récapitulatif des lois usuelles :
  – Loi de **Bernoulli** de paramètre $p \in [0, 1]$:

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p.$$

  – Loi **binomiale** $\mathcal{B}(n, p)$ de paramètre $(n, p) \in \mathbb{N} \times [0, 1]$:

$$\mathbb{P}(X = k) = C_n^k\, p^k(1 - p)^{n-k}, \quad k \in \{0, \ldots, n\}.$$

  – Loi **géométrique** de paramètre $p \in]0, 1]$:

$$\mathbb{P}(X = k) = p(1 - p)^{k-1}, \quad k \in \mathbb{N}^*.$$

  – Loi de **Poisson** de paramètre $\theta \in]0, \infty[$:

$$\mathbb{P}(X = k) = \frac{\theta^k}{k!}\,\mathrm{e}^{-\theta}, \quad k \in \mathbb{N}.$$

Loi	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	$\phi_X(z)$
Bernoulli $p \in [0, 1]$	p	$p(1-p)$	$1 - p + pz$
binomiale $(n, p) \in \mathbb{N} \times [0, 1]$	np	$np(1-p)$	$(1 - p + pz)^n$
géométrique $p \in]0, 1]$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pz}{1-(1-p)z}$
Poisson $\theta \in]0, \infty[$	θ	θ	$\mathrm{e}^{-\theta(1-z)}$

II.17 Exercices

Les exercices dans la partie du cours sont aux pages suivantes

Exercice II.1 p. 31, *Exercice* II.2 p. 32, *Exercice* II.3 p. 38, *Exercice* II.4 p. 39, *Exercice* II.5 p. 46, *Exercice* II.6 p. 46, *Exercice* II.7 p. 50, *Exercice* II.8 p. 50, *Exercice* II.9 p. 53, *Exercice* II.10 p. 57, *Exercice* II.11 p. 58, *Exercice* II.12 p. 60, *Exercice* II.13 p. 60, *Exercice* II.14 p. 63.

Exercice II.15.
On effectue une enquête sur la qualité de N=100 hôpitaux qui ont tous pratiqué n=10 opérations de l'appendicite. On sait que le taux de réussite de cette opération est de $\tau = 97,5\%$.

1. Sur l'année 1998 un hôpital a eu 3 échecs. Peut-on dire que l'hôpital est mauvais ?
2. Quelle est la probabilité pour que ce même l'hôpital ait 3 échecs en 1999 ? Conclusion.
3. Un hôpital a eu 3 échecs en 1997 et 3 échecs en 1998. Que peut-on dire de cet hôpital ?

△

Exercice II.16.
Trouver des v.a.d. positives (X, Y) telles que a) $\mathbb{E}[X] > \mathbb{E}[Y]$ et $\mathbb{P}(X > Y) \geq 1/2$, puis telles que b) $\mathbb{E}[X] < \mathbb{E}[Y]$ et $\mathbb{P}(X > Y) \geq 1/2$. △

Exercice II.17.
Problème du chevalier de Méré (1600). A-t-on plus de chance d'obtenir au moins un six en lançant un dé équilibré quatre fois de suite ou un double six en lançant deux dés équilibrés vingt-quatre fois de suite ?

1. On note X_1 le nombre de six obtenu lorsqu'on lance le dé quatre fois de suite, et X_2 le nombre de double six lorsqu'on lance deux dés vingt-quatre fois de suite. Pour répondre à la question du chevalier de Méré, comparer $\mathbb{P}(X_1 \geq 1)$ et $\mathbb{P}(X_2 \geq 1)$.
2. Calculer et comparer $\mathbb{E}[X_1]$ et $\mathbb{E}[X_2]$.

△

Exercice II.18.
Inégalité de Markov. En s'inspirant de l'inégalité de Tchebychev, montrer que

$\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}[|X|]}{a}$ pour toute v.a. réelle X. Montrer plus généralement que si h est une fonction croissante positive mesurable, alors $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[h(X)]}{h(a)}$. △

Exercice II.19.
Inégalité de Jensen. Soit φ une fonction réelle convexe : pour tout $x, y \in \mathbb{R}$, et $t \in [0,1]$, $\varphi(tx + (1-t)y) \geq t\varphi(x) + (1-t)\varphi(y)$. Il existe une autre caractérisation des fonctions convexes : pour tout $a \in \mathbb{R}$, il existe $\lambda_a \in \mathbb{R}$ tel que $\varphi(a) + \lambda_a(x-a) \leq \varphi(x)$.

1. On suppose que X est une variable aléatoire réelle intégrable. Montrer que $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$, dès que $\mathbb{E}[\varphi(X)]$ a un sens.
2. En déduire que si X est intégrable, alors $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$.
3. Montrer que pour toute suite de réels positifs $a_1, \ldots, a_n$, on a :
$$n^{-1}\left(\sum_{i=1}^{n} a_i\right)^2 \leq \sum_{i=1}^{n} a_i^2.$$
4. Caractérisez les cas où l'égalité est vérifiée. (On rappelle que si $\mathrm{Var}(Y) = 0$ alors Y est constante p.s.)

△

Exercice II.20.
Soit $(X_1, \ldots, X_n)$ des variables aléatoires discrètes indépendantes. Montrer que $(X_1 + \cdots + X_{p_1}), \ldots, (X_{p_{k-1}+1} + \cdots + X_{p_k})$ où $1 \leq p_1 < \cdots < p_k = n$, sont indépendantes. △

Exercice II.21.
Soit T_1 et T_2 deux variables aléatoires indépendantes de loi géométrique de paramètre p_1 et p_2.

1. Calculer et reconnaître la loi de $\min(T_1, T_2)$.
2. Calculer la loi jointe de $\min(T_1, T_2)$ et $T_1 - T_2$.
3. En déduire que $\min(T_1, T_2)$ est indépendant de $\mathbf{1}_{\{T_1 \leq T_2\}}$. Quelle est la loi de $\mathbf{1}_{\{T_1 \leq T_2\}}$?
4. Déduire également de la question 2 que $R = \max(T_1, T_2) - \min(T_1, T_2)$ est indépendant de $\min(T_1, T_2)$.
5. Calculer la loi de R conditionnellement à $\{R \neq 0\}$. Reconnaître cette loi quand $p_1 = p_2$.

△

Exercice II.22.
Soit X une v.a. géométrique de paramètre p, et Z une v.a.d. à valeurs entières, indépendante de X. Montrer que $\mathbb{P}(X > Z) = \phi_Z(1-p)$. △

Exercice II.23.
Autre méthode pour le calcul de l'espérance d'une v.a. à valeurs entières.

1. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Montrer que :
$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$
2. Montrer la formule suivante par récurrence : pour $a \geq 1$,
$$\sum_{k=0}^{n} \mathrm{C}_{k+a-1}^{a-1} = \mathrm{C}_{n+a}^{a}.$$
3. On considère une urne contenant b boules bleues et r boules rouges. On note X_b le nombre de boules qu'il faut sortir de l'urne avant de voir apparaître une boule bleue. Calculer $\mathbb{E}[X_b]$ en utilisant les questions précédentes.

△

Exercice II.24.
On désire modéliser le temps d'attente d'une panne de machine à l'aide d'une variable aléatoire sans mémoire : la probabilité pour que la machine tombe en panne après la date $k+n$ sachant qu'elle fonctionne à l'instant n est indépendante de n.

1. Montrer que la loi géométrique de paramètre p est sans mémoire c'est-à-dire que $\mathbb{P}(X > k+n | X > n)$ est indépendant de n.
2. Caractériser toutes les lois des variables aléatoires X à valeurs dans $\mathbb{N}^*$ qui sont sans mémoire. On pourra calculer $\mathbb{P}(X > 1+n)$ en fonction de $\mathbb{P}(X > 0)$.
3. Caractériser toutes les lois des variables aléatoires X à valeurs dans $\mathbb{N}$ qui sont sans mémoire.

△

Exercice II.25.
Autour de la loi binomiale et la loi de Poisson.

1. Déterminer la loi de $X_1 + X_2$ où X_1 et X_2 sont des variables indépendantes de loi binomiale de paramètre respectif (n_1, p) et (n_2, p). (On pourra considérer l'égalité suivante : $(1+x)^{n_1}(1+x)^{n_2} = (1+x)^{n_1+n_2}$.)

2. Retrouver ce résultat en interprétant la loi binomiale comme la loi de la somme de variables aléatoires de Bernoulli indépendantes.
3. Déterminer la loi de $Y_1 + Y_2$ où Y_1 et Y_2 sont des variables indépendantes de Poisson de paramètre respectif θ_1 et θ_2. Comment pouviez-vous deviner le résultat ?
4. Calculer la loi de X_1 sachant que $X_1 + X_2 = n$.
5. Calculer la limite de $\mathbb{P}(X_1 = k | X_1 + X_2 = n)$ quand $n_1 \to \infty$, $n_2 \to \infty$, $p \to 0$ et $n_1 p = \theta_1$, $n_2 p = \theta_2$.
6. Déterminer la loi de Y_1 sachant que $Y_1 + Y_2 = n$. Comment pouviez-vous avoir une intuition du résultat ?

△

Exercice II.26.
On considère une urne contenant r boules rouges et b boules bleues. On tire au hasard les boules **sans remise**.

1. Combien existe-t-il de tirages complets possibles ?
2. On note X_n le nombre de boules rouges obtenues alors que l'on a tiré n boules. Calculer la loi de X_n.
3. Reconnaître la loi limite de X_n quand $r \to \infty$ et $r/(r+b) \to p \in]0,1[$.
4. On note $Y_k = 1$ si la k-ième boule est rouge et $Y_k = 0$ sinon. Quelle est la loi de $(Y_1, \ldots, Y_{r+b})$?
5. En déduire que $Y_1, \ldots, Y_{r+b}$ ont même loi. Calculer la loi de Y_1.
6. Exprimer X_n en fonction de $Y_1, \ldots, Y_n$. Calculer $\mathbb{E}[X_n]$ et $\mathrm{Var}(X_n)$.
7. On note S_n le nombre de boules rouges lors d'un tirage aléatoire de n boules de l'urne **avec remise**. Quelle est la loi de S_n ? Comparer avec la question 3.
8. Calculer $\mathbb{E}[S_n]$ et $\mathrm{Var}(S_n)$. Comparer avec la question 6.

△

Exercice II.27.
On considère une urne contenant r boules rouges et b boules bleues. On tire au hasard les boules **sans remise**.

1. On note T_1 le nombre de boules qu'il faut tirer pour obtenir une boule rouge. Calculer la loi de T_1.
2. Quelle est la loi limite quand $r \to \infty$ et $r/(r+b) \to p \in]0,1[$. Comparer la loi limite avec le premier temps d'obtention d'une boule rouge dans un tirage **avec remise**.

3. On note $Z_1 = T_1 - 1$ et pour $i \in \{2, \dots, r\}$, Z_k le nombre de boules bleues obtenues entre la $(k-1)$-ième boule rouge et la k-ième boule rouge. Enfin, on note Z_{r+1} le nombre de boules bleues obtenues après la dernière boule rouge. Calculer la loi de $(Z_1, \dots, Z_{r+1})$.
4. En déduire que $Z_1, \dots, Z_{r+1}$ ont même loi. Calculer $Z_1 + \cdots + Z_{r+1}$, puis calculer $\mathbb{E}[T_1]$. Cette méthode ne permet pas néanmoins de calculer $\mathrm{Var}(T_1)$. On admet que $\mathrm{Var}(T_1) = \dfrac{rb(b+r+1)}{(r+1)^2(r+2)}$.
5. Que se passe-t-il pour $r \to \infty$ et $r/(r+b) \to p \in]0,1[$?

△

Exercice II.28.
On considère deux urnes contenant chacune b boules bleues et r boules rouges. On note X le nombre de fois où quand on retire une boule de chacune des deux urnes (sans remise), les deux boules sont de la même couleur.

1. Calculer la loi de X.
2. La loi explicite de X ne permet pas de calculer facilement son espérance ou sa variance. Pour cela on introduit les variables $Y_1, \dots, Y_n$ où $Y_i = 1$ si les $i^{\text{ème}}$ boules des deux urnes sont de la même couleur. Calculer la loi de $Y_1, \dots, Y_{b+r}$.
3. En déduire que les variables $Y_1, \dots, Y_{b+r}$ ont même loi.
4. Calculer $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

△

Exercice II.29.
Votre petit frère collectionne les images des joueurs de la coupe du monde que l'on trouve dans les tablettes de chocolat. On suppose qu'il existe N images différentes et qu'elles sont équitablement réparties, à raison de une par tablette. On note X_r le nombre de tablettes qu'il faut acheter pour avoir r images différentes. On note T_n le nombre de tablettes qu'il faut acheter pour avoir une nouvelle image sachant que l'on en a déjà $n-1$. ($T_1 = 1$.)

1. Montrer que la loi de T_2 est une loi géométrique dont on donnera le paramètre.
2. Donner la loi de T_n.
3. En déduire $\mathbb{E}[X_r]$ et $\mathbb{E}[X_N]$. Donner un équivalent de $\mathbb{E}[X_N]$ quand $N \to \infty$.
4. On admet que les v.a.d. T_n sont indépendantes. Calculer $\mathrm{Var}(X_N)$. En donner un équivalent quand $N \to \infty$.
5. Votre petit frère est uniquement intéressé par l'équipe française (k joueurs). On note Y_k le nombre de tablettes qu'il faut acheter pour obtenir les k joueurs français. Calculer $\mathbb{E}[Y_k]$ et $\mathrm{Var}(Y_k)$.

△

Exercice II.30.
Deux joueurs A et B jouent à pile ou face. La mise est de 1 Euro. La probabilité que A gagne la mise est $p \in]0,1[$. A possède a Euro et B, b Euro. Le jeu s'arrête quand l'un des deux joueurs est ruiné. On note T l'instant où le jeu s'arrête.

1. Montrer que presque sûrement T est fini. (On pourra comparer T avec le temps d'attente de $b+a$ piles successifs.)
2. On note R l'évènement suivant : A est ruiné avant B. On désire calculer la probabilité de R. On note $h(x)$ la probabilité que A soit ruiné avant qu'il ne possède $a+b$, alors qu'il possède maintenant x. Ainsi $h(a) = \mathbb{P}(R)$. Calculer $h(0)$ et $h(a+b)$.
3. Montrer que : pour $0 < x < a+b$,
$$h(x) = ph(x+1) + (1-p)h(x-1).$$
4. En déduire que :
$$h(x+1) - h(x) = p^{-1}(1-p)[h(x) - h(x-1)], \quad 0 < x < a+b.$$
5. Montrer que si $p = 1/2$, alors $\mathbb{P}(R) = b/(a+b)$.
6. Montrer que si $p \neq 1/2$, alors $h(x) = \dfrac{\rho^{a+b} - \rho^x}{\rho^{a+b} - 1}$, où $\rho = (1-p)/p$. En déduire que $\mathbb{P}(R) = (\rho^{a+b} - \rho^a)/(\rho^{a+b} - 1)$. Quelle est la limite de $\mathbb{P}(R)$ quand $p \to 1/2$?

△

Exercice II.31.
On considère un jeu de pile ou face biaisé : les variables aléatoires $(X_n, n \in \mathbb{N}^*)$ sont indépendantes et de même loi de Bernoulli de paramètre $p \in]0,1[$: $\mathbb{P}(X_n = 1) = p$ et $\mathbb{P}(X_n = 0) = 1-p$. On note T_k l'instant du k-ième succès : $T_1 = \inf\{n \geq 1; X_n = 1\}$ et : pour $k \geq 2$,
$$T_k = \inf\{n \geq T_{k-1} + 1; X_n = 1\}.$$

1. Montrer que T_1 et $T_2 - T_1$ sont indépendants.
2. On pose $T_0 = 0$. Montrer que $T_1 - T_0, T_2 - T_1, \cdots, T_{k+1} - T_k$ sont indépendants et de même loi.
3. Calculer $\mathbb{E}[T_k]$ et $\mathrm{Var}(T_k)$.
4. Déterminer $\mathbb{P}(T_k = n)$ directement. Donner la fonction génératrice de T_k. La loi de T_k est appelée loi binomiale négative de paramètre (k,p).

On possède une seconde pièce de paramètre $\rho \in]0,1[$. On note τ l'instant du premier succès de la seconde pièce. On décide de jouer avec la première pièce jusqu'au τ-ième succès (c'est-à-dire T_τ).

5. Déterminer la loi de T_τ à l'aide des fonctions génératrices. Reconnaître la loi de T_τ.
6. Retrouver ce résultat à l'aide d'un raisonnement probabiliste sur les premiers temps de succès.

△

Exercice II.32.
Un singe est devant une machine à écrire (26 touches correspondant aux 26 caractères). Il frappe sur les touches au hasard.

1. Quelle est la probabilité pour qu'il frappe la lettre A en un temps fini ?
2. Quelle est la probabilité pour qu'il frappe correctement en un temps fini "VICTORHUGO" (resp. l'œuvre complète de Victor Hugo) ? En introduisant une variable géométrique de paramètre 26^{-10}, donner une majoration de l'espérance du temps de première occurrence du mot VICTORHUGO.

△

Exercice II.33.
Le calcul exact du temps moyen de première occurrence d'une séquence est délicat. Mais on peut l'expliciter sur un modèle simple. Pour cela on considère une suite de variables de Bernoulli indépendantes et identiquement distribuées $(X_i, i \in \mathbb{N}^*)$. On note $p = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = 1 - q$. On note T_{11} le premier temps d'occurrence de la séquence 11 :

$$T_{11} = \inf \{k \geq 2; X_{k-1} = 1, X_k = 1\}.$$

De manière similaire, on note T_0, T_1 et T_{10} les temps respectifs d'apparition des séquences 0, 1 et 10.

1. Rappeler les lois de T_0 et T_1.
2. On note $\pi_k = \mathbb{P}(T_{11} = k)$. Calculer π_k pour $k \in \{1,2,3,4\}$. Montrer que : pour $k \geq 3$,
$$\pi_k = q\pi_{k-1} + pq\pi_{k-2}.$$
3. En déduire la fonction génératrice de T_{11}. Calculer $\mathbb{E}[T_{11}]$.
4. Calculer $\mathbb{P}(T_{10} = k, T_1 = n)$. Donner la fonction génératrice de T_{10}. En déduire que T_{10} a même loi que la somme de deux variables géométriques indépendantes dont on précisera les paramètres. Calculer $\mathbb{E}[T_{10}]$.

5. Pour $p = 1/2$, calculer $\mathbb{E}[T_{11}]$, $\mathbb{E}[T_{10}]$ et $\mathbb{P}(T_{10} > T_{11})$. Commentaires.

△

Exercice II.34.
On considère un central téléphonique d'une entreprise de vente par correspondance. Il y a un grand nombre de communications téléphoniques par jour en France. Un faible pourcentage est destiné à l'entreprise de VPC. On modélise donc le nombre N de communications téléphoniques que reçoit l'entreprise par jour à l'aide d'une variable aléatoire de Poisson de paramètre θ. L'entreprise possède un service de vente et un service après-vente. Chaque communication a une probabilité p de concerner le service de vente.

1. Calculer, à l'aide des fonctions génératrices, la loi de N_V du nombre de communications par jour que reçoit le service de vente.
2. Calculer la loi de (N_V, N_A), où N_A est le nombre de communications par jour que reçoit le service après-vente.
3. En déduire la loi de la somme de deux variables de Poisson indépendantes de paramètre θ_1 et θ_2.

△

Exercice II.35.
On souhaite répondre à la question suivante : Existe-t-il une stratégie gagnante à horizon fini pour un jeu de pile ou face équilibré ?

À l'étape n, on lance une pièce non biaisée. Si on obtient pile, on gagne 1 Euro, sinon on perd 1 Euro. On note $X_n \in \{-1, 1\}$ le gain obtenu à l'étape n. La richesse à l'instant n est $S_n = \sum_{i=1}^{n} X_i$, la richesse initiale étant nulle. Les v.a.d. $(X_i, i \in \mathbb{N}^*)$ sont indépendantes et de même loi : $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. Si on s'arrête à l'instant N, on gagne en moyenne $\mathbb{E}[S_N] = 0$.

Une stratégie est une règle de décision qui dit à l'instant n, en fonction des résultats jusqu'à cet instant, si on joue le coup suivant ou non. Plus précisément, on se donne au début du jeu une suite de fonctions $(f_k, k \in \{1, \dots, N\})$ où f_k est définie sur $\{-1, 1\}^k$ (cet ensemble correspond à l'ensemble des résultats possibles des lancers jusqu'au k-ième lancer) à valeurs dans $\{0, 1\}$. Si $f_n(x_1, \dots, x_n) = 1$, cela signifie que si on a lancé n fois la pièce et que l'on a observé la séquence $(x_1, \dots, x_n)$, alors on joue le $n+1$-ième coup. Si $f_n(x_1, \dots, x_n) = 0$, cela signifie que soit on s'est déjà arrêté avant l'instant n, soit on s'arrête de jouer à l'instant n. Si τ est le premier instant où l'on s'arrête, alors $\mathbf{1}_{\{\tau > n\}} = f_n(X_1, \dots, X_n)$.

On suppose que l'horizon du jeu est fini. On joue au plus N coups : i.e. $f_N = 0$. Calculer $\mathbb{E}[S_\tau]$, en décomposant suivant les valeurs de τ puis en faisant intervenir les fonctions f_k. Commentaires. △

Exercice II.36.
Théorème de Weierstrass (1885) : "Toute fonction continue sur un intervalle fermé borné est limite uniforme d'une suite de polynômes".

Cet exercice s'inspire de la démonstration de Bernstein du théorème de Weierstrass. Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $x \in [0,1]$. Pour $n \geq 1$, on considère la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Soit $h : [0,1] \to \mathbb{R}$ une fonction continue. Soit $\delta > 0$. On pose $\Delta_n = \{\left|\bar{X}_n - x\right| > \delta\}$.

1. Montrer que $\mathbb{P}(\Delta_n) \leq \delta^{-2}\mathbb{E}[(\bar{X}_n - x)^2]$. Majorer $\mathbb{P}(\Delta_n)$ indépendamment de $x \in [0,1]$.
2. Déterminer $\lim\limits_{n\to\infty} \sup\limits_{x\in[0,1]} \left|h(x) - \mathbb{E}[h(\bar{X}_n)]\right|$, en écrivant :

$$\left|h(x) - h(\bar{X}_n)\right| = \left|h(x) - h(\bar{X}_n)\right| \mathbf{1}_{\Delta_n} + \left|h(x) - h(\bar{X}_n)\right| \mathbf{1}_{\Delta_n^c}.$$

3. Quelle est la loi de $n\bar{X}_n$?
4. En déduire que :

$$\lim_{n\to\infty} \sup_{x\in[0,1]} \left| h(x) - \sum_{k=0}^n \binom{n}{k} h(k/n) x^k (1-x)^{n-k} \right| = 0.$$

5. Soit $f : \mathbb{R}^+ \to \mathbb{R}$ continue bornée. Montrer, en s'inspirant des questions précédentes, que : pour tout $x \in \mathbb{R}^+$,

$$\lim_{n\to\infty} \left| f(x) - \sum_{k=0}^\infty \mathrm{e}^{-nx} \frac{(nx)^k}{k!} f(k/n) \right| = 0.$$

Si l'on suppose f uniformément continue, la convergence ci-dessus est-elle uniforme en x ? (Prendre par exemple $f(x) = \cos(x)$ pour $x_n = \pi n$.)

△

Exercice II.37.
Optimisation de coûts. Le coût de dépistage de la maladie M à l'aide d'un test sanguin est c. La probabilité qu'une personne soit atteinte de la maladie M est p. Pour effectuer un dépistage parmi N personnes, on propose les deux méthodes suivantes :

– Un test par personne.
– On mélange les prélèvements sanguins de n personnes et on effectue le test. Si on détecte la maladie M, alors on refait un test sanguin pour chacune des n personnes. Calculer le coût moyen de cette stratégie. On supposera $np \ll 1$, et on montrera que $n \simeq p^{-1/2}$ est une taille qui minimise correctement le coût du dépistage.

Quelle méthode choisissez-vous ? △

Exercice II.38.
On désire répondre à la question suivante : Peut-on reproduire le résultat d'un lancer d'un dé équilibré à onze faces, numérotées de 2 à 12, comme la somme d'un lancer de deux dés à six faces, numérotées de 1 à 6, éventuellement différemment biaisés ?

1. Soit X de loi uniforme sur $\{2, \dots, 12\}$. Vérifier que la fonction génératrice de X est un polynôme. Quelles sont ses racines réelles ?
2. Étudier les racines de la fonction génératrice associée à la somme d'un lancer de deux dés à six faces. Conclure.

△

Exercice II.39.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in]0,1[$. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. La loi faible des grands nombres assure que la suite $(\bar{X}_n, n \in \mathbb{N}^*)$ converge en probabilité vers $p = \mathbb{E}[X_1]$. On s'intéresse aux évènements rares du type $\{\bar{X}_n > a\}$ ou $\{\bar{X}_n < b\}$ où $0 < b < p < a < 1$. En posant $\bar{Y}_n = 1 - \bar{X}_n$, on remarque qu'il suffit d'étudier l'évènement $\{\bar{X}_n > a\}$ pour $p < a < 1$.

1. Montrer que pour $\lambda > 0$, on a $\mathbb{P}(\bar{X}_n > a) \leq \mathbb{E}\left[\mathrm{e}^{\lambda X_1}\right]^n \mathrm{e}^{-an\lambda}$.

On considère la transformée de Cramer de la loi de X_1, Λ_p définie sur $[0,1]$ par :

$$\Lambda_p(x) = x \log\Big(x/p\Big) + (1-x)\log\Big((1-x)/(1-p)\Big).$$

2. Montrer que $\mathbb{P}(\bar{X}_n > a) \leq \mathrm{e}^{-n\Lambda_p(a)}$.
3. Montrer que la fonction $\Lambda_p(x)$ atteint son unique minimum en $x = p$.

On définit l'entropie de la loi de Bernoulli de paramètre p par $H_p = -p\log p - q\log q$, où $q = 1-p$.

4. Pour quelle valeur de p l'entropie est-elle maximale ?

On considère Ω l'ensemble des suites $\omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n$ muni de la probabilité $\mathbb{P}(\{\omega\}) = p^{\sum_{i=1}^n \omega_i} q^{n - \sum_{i=1}^n \omega_i}$. On considère le sous-ensemble de Ω des suites typiques de longueur n défini par :

$$C_n = \Big\{\omega; \Big| \frac{1}{n}\sum_{i=1}^{n} \omega_i - p\Big| \leq \delta_n\Big\},$$

où $\lim_{n\to\infty} \delta_n = 0$ et $\lim_{n\to\infty} \sqrt{n}\delta_n = +\infty$.

5. Soit $\alpha \in]0,1[$. Montrer, à l'aide de la question 2, que pour n assez grand, on a $\mathbb{P}(C_n) \geq 1 - \alpha$.
6. Montrer que pour n assez grand, on a : pour tout $\omega \in C_n$,

$$\mathrm{e}^{-n(H_p+\beta_n)} \leq \mathbb{P}(\{\omega\}) \leq \mathrm{e}^{-n(H_p-\beta_n/2)} \leq \mathrm{e}^{-n(H_p-\beta_n)},$$

   avec $0 \leq \beta_n = c_p\delta_n$, la constante c_p dépendant seulement de p.
7. Montrer que pour tout n assez grand, on a :

$$\mathrm{e}^{n(H_p-\beta_n)} \leq \mathrm{Card}\ (C_n) \leq \mathrm{e}^{n(H_p+\beta_n)},$$

8. Quel résultat obtenez-vous si $p = 1/2$, si $p \simeq 0$ ou $p \simeq 1$?

Certaines techniques de compression consistent à trouver les suites typiques pour une longueur n fixée, et à les coder par des séquences courtes (d'où la compression). La contrepartie est que les suites non-typiques sont codées par des séquences plus longues. Comme les suites non-typiques sont très rares, elles apparaissent peu souvent et donc la compression est peu dégradée par les suites non-typiques. La compression est d'autant plus importante que l'ensemble des suites typiques est petit. Dans le cas de séquences aléatoires traitées dans cet exercice, cela correspond aux cas $p \simeq 0$ ou $p \simeq 1$.

△

III

Variables aléatoires à densité

Ce chapitre est consacré à l'étude des variables aléatoires à densité qui sont d'un usage très fréquent en modélisation. Intuitivement, la probabilité qu'une variable aléatoire de densité f prenne une valeur dans l'intervalle infinitésimal $[x, x+dx]$ est $f(x)\ dx$. Contrairement aux variables aléatoires discrètes, les variables aléatoires à densité prennent un continuum de valeurs possibles.

On donne dans le paragraphe III.1 la définition des variables aléatoires à densité, ainsi que la définition de la fonction de répartition d'une variable aléatoire quelconque. On admet une formule explicite pour le calcul de l'espérance d'une fonction d'une variable aléatoire à densité au paragraphe III.3. Les paragraphes III.4 et III.5 donnent plusieurs exemples de variables aléatoires à densité qui sont fréquemment utilisées.

On présente une caractérisation de l'indépendance pour des variables aléatoires à densité au paragraphe III.6. On détaille dans les paragraphes III.2 et III.7 des méthodes pour calculer des lois de variables aléatoires à densité à partir d'autres variables aléatoires à densité. Les notions de loi conditionnelle et d'espérance conditionnelle sont présentées au paragraphe III.8.

Le paragraphe III.9 est consacré aux méthodes de simulations de variables aléatoires usuelles à partir de générateurs de nombres (pseudo-aléatoires).

Enfin le paragraphe III.10 rappelle quelques résultats de l'intégration par rapport à la mesure de Lebesgue.

III.1 Définitions

Vous roulez à vélo, et votre pneu crève. Intuitivement, la probabilité pour que la crevaison ait lieu sur l'arc $\overset{\frown}{AB}$ est proportionnelle à sa longueur : $\theta/2\pi$, où θ est l'angle correspondant à l'arc $\overset{\frown}{AB}$. La position de la crevaison est uniformément répartie sur la roue. Cet exemple ne peut pas être modélisé à l'aide de variables aléatoires discrètes.

Définition III.1. *On dit que la loi d'une variable aléatoire réelle X est à densité s'il existe une fonction f positive mesurable de $\mathbb{R}$ dans $[0,+\infty]$ qui vérifie $\int_{\mathbb{R}} f(x)\,dx = 1$ et : pour tout ensemble $A \in \mathcal{B}(\mathbb{R})$,*

$$\boxed{\mathbb{P}(X \in A) = \int_A f(x)\,dx.}$$

*On dit que f, parfois notée f_X, est la **densité** de la loi de X. On identifie la **loi** de X notée $\mathcal{L}(X)$ à sa densité f_X.*

Exemple. Dans le cas de la crevaison, on note X l'angle entre la crevaison et la valve. La densité de la loi de X est $f(x) = \dfrac{1}{2\pi}\,\mathbf{1}_{[0,2\pi]}(x)$. On dit que la loi de X est la loi **uniforme** sur $[0, 2\pi]$. ◊

Remarque. Soit $a \leq b$. On a d'après la définition $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$. En particulier si $a = b$, on a $\mathbb{P}(X = a) = \int_a^a f(y)dy = 0$. On démontre en utilisant la σ-additivité que pour tout $A \subset \mathbb{R}$ au plus dénombrable, on a $\mathbb{P}(X \in A) = 0$. Ce comportement est différent de celui des v.a. discrètes. ◊

Remarque III.2. La densité est-elle unique ? On rappelle (ou admet) que si deux fonctions intégrables satisfont $\int_a^b f(x)dx = \int_a^b g(x)dx$ pour tous $a \leq b$, alors pour tout borélien A, $\int_A f(x)dx = \int_A g(x)dx$. De plus l'ensemble $\{x; f(x) \neq g(x)\}$ (mesurable) est de mesure nulle pour la mesure de Lebesgue. On dit que $f = g$ presque partout. L'égalité presque partout définit en fait une relation d'équivalence dans l'ensemble $\mathcal{L}^1$ des fonctions mesurables intégrables. On note souvent L^1 l'ensemble $\mathcal{L}^1$ quotienté par cette relation d'équivalence. On a l'unicité dans L^1 de la densité de la loi de X. Ainsi les fonctions $\dfrac{1}{2\pi}\,\mathbf{1}_{[0,2\pi]}(x)$ et $\dfrac{1}{2\pi}\,\mathbf{1}_{]0,2\pi[}(x)$ sont égales presque partout et, dans l'exemple ci-dessus, elles définissent le même représentant dans L^1 de la densité. ◊

Définition III.3. *Soit X une v.a. réelle. La* ***fonction de répartition*** *de X est la fonction mesurable F, parfois notée F_X, définie sur $\mathbb{R}$ par :*

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Si X est une variable aléatoire discrète, alors sa fonction de répartition est constante par morceaux. Si la loi de X possède une densité continue f, alors $F(x) = \int_{-\infty}^{x} f(y)\, dy$ pour tout $x \in \mathbb{R}$. En particulier, si la densité f est continue, la fonction de répartition est continue et même de classe $\mathcal{C}^1$ et de dérivée f.

La figure III.1 présente plusieurs fonctions de répartition : loi de Poisson, loi uniforme, loi gaussienne (voir paragraphe III.4) et loi du temps d'attente à un feu tricolore. Si le feu est vert, le temps d'attente est nul ; si le feu est rouge, le temps d'attente correspond à une variable aléatoire de loi uniforme sur $[0, T]$, où T est la durée du cycle rouge. La fonction de répartition du temps d'attente présente un saut (feu vert) et une partie linéaire croissante (feu rouge). Le temps d'attente au feu rouge n'est pas une v.a. discrète et sa loi ne possède pas non plus de densité.

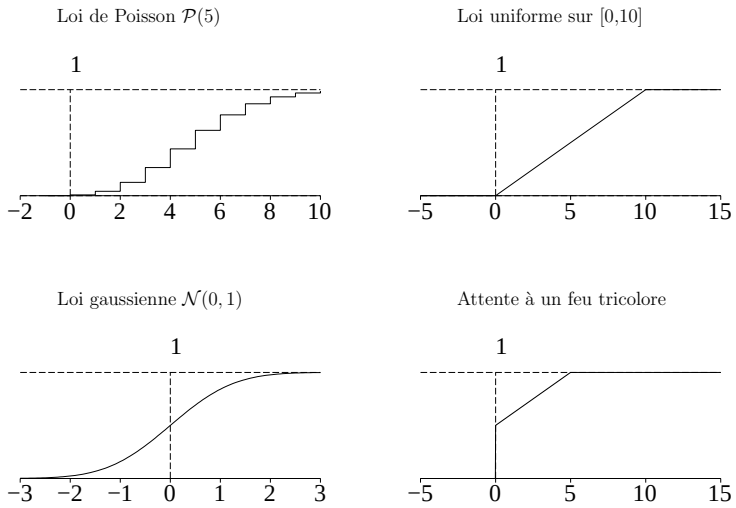


Figure III.1. Quelques fonctions de répartition.

Remarque. On peut montrer que pour toute fonction croissante F, continue à droite, telle que $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$, alors il existe un espace probabilisé et une variable aléatoire X, définie sur cet espace, telle que F est sa

fonction de répartition. Il est facile de vérifier cette assertion dans le cas particulier où il existe $f > 0$ telle que $F(x) = \int_{-\infty}^{x} f(y)\, dy$ pour tout $x \in \mathbb{R}$. En effet, dans ce cas la fonction F est une bijection de $\mathbb{R}$ dans $]0,1[$. Elle est continue ainsi que sa réciproque. Si on pose $\Omega =]0,1[$, $\mathcal{F}$ la tribu borélienne sur $]0,1[$, et $\mathbb{P}$ la mesure de Lebesgue sur $]0,1[$, alors $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé. On l'a déjà utilisé lors de la construction du schéma de Bernoulli au paragraphe II.5. On définit la fonction mesurable X de $]0,1[$ dans $\mathbb{R}$ par $X(\omega) = F^{-1}(\omega)$, $\omega \in \Omega$. On a ainsi :

$$\mathbb{P}(X \in [a,b]) = \mathbb{P}(\omega \in [F^{-1}(a), F^{-1}(b)]) = \int_{F^{-1}(a)}^{F^{-1}(b)} dx = \int_a^b f(y)\, dy,$$

en utilisant le changement de variable $x = F(y)$ pour la dernière égalité. Comme ceci est valable pour tout $a, b \in \mathbb{R}$, on déduit de la définition III.1 et du théorème III.5 que f est la densité de la loi de X et F sa fonction de répartition. On peut généraliser cette construction à F quelconque. ◊

On peut généraliser la notion de densité et de fonction de répartition aux v.a. vectorielles. On renvoie au chapitre III.10 pour la définition des intégrales multiples.

Définition III.4. *Soit $X = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$.*

- *La loi de X est à densité s'il existe une fonction f, appelée densité de la loi de X et parfois notée f_X, positive mesurable de $\mathbb{R}^d$ dans $[0, +\infty]$ qui vérifie $\int_{\mathbb{R}^d} f(x)\, dx = 1$ et : pour tout ensemble $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\boxed{\mathbb{P}(X \in A) = \int_A f(x)\, dx.} \tag{III.1}$$

- *La fonction de répartition de X est la fonction mesurable F, parfois notée F_X, définie sur $\mathbb{R}^d$ par :*

$$F(x) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Par abus de langage, on dit qu'une variable aléatoire dont la loi possède une densité est une **variable aléatoire à densité** ou une **variable aléatoire continue** (v.a.c.). (On signale que dans certains ouvrages une variable aléatoire réelle ou vectorielle est dite continue si sa fonction de répartition est continue ; elle est dite absolument continue si sa loi possède une densité.)

On admet le théorème important suivant.

Théorème III.5. *La fonction de répartition caractérise la loi : deux v.a. (réelle ou vectorielle) ont même loi si et seulement si elles ont même fonction de répartition.*

La notion de quantile pour des v.a. réelles sera utilisée en statistique.

Définition III.6. *Soit Y une variable aléatoire réelle. Le* ***quantile*** *(on parle aussi de fractile), $a_r \in \mathbb{R}$, d'ordre $r \in]0,1[$ de la loi de Y est défini par :*

$$\boxed{a_r = \inf\{x; F(x) \geq r\}.} \tag{III.2}$$

La fonction $F^{-1} : r \mapsto a_r$ est l'inverse généralisé continu à gauche de la fonction F. Les propriétés des inverses généralisés impliquent la proposition suivante qui est admise.

Proposition III.7. *On reprend les notations de la définition III.6. On a $F(a_r) \geq r$. Si la fonction de répartition de Y est continue, alors $F(a_r) = r$. Si la fonction de répartition est de plus strictement croissante au point a_r, alors a_r est l'unique solution de l'équation $F(x) = r$.*

III.2 Lois marginales

Si $X = (X_1, \ldots, X_d)$ est une v.a.c. à valeurs dans $\mathbb{R}^d$, la i-ème composante X_i du vecteur est une v.a.c. réelle. On peut, comme dans le cas des v.a. discrètes, retrouver la loi marginale de X_i quand on connaît la loi du vecteur X. On calcule les lois marginales de X. Quitte à permuter i et 1, il suffit de calculer la loi de X_1.

Proposition III.8. Formule des lois marginales. *La v.a. X_1 est une v.a. continue. De plus la densité de sa loi est donnée par : pour $x \in \mathbb{R}$,*

$$\boxed{f_{X_1}(x) = \int_{\mathbb{R}^{d-1}} dx_2 \cdots dx_d \; f_X(x, x_2, \ldots, x_d).}$$

Cette expression est à rapprocher de la formule des lois marginales pour les v.a.d. (proposition II.9).

Démonstration. Soit Π_1 la projection de $\mathbb{R}^d$ dans $\mathbb{R}$ qui à $(x_1, \dots, x_d)$ associe x_1. Cette application est continue car l'image réciproque de tout ouvert est un ouvert. Grâce au corollaire I.18, elle est donc mesurable. On a $X_1 = \Pi_1 \circ X$. L'application X_1 est la composée de deux applications mesurables, elle est donc mesurable. Elle est clairement à valeurs réelles.

La formule des lois marginales se déduit de la définition. En effet, si $A \in \mathcal{B}(\mathbb{R})$, alors, on a grâce à (III.1) :

$$\begin{aligned}\mathbb{P}(X_1 \in A) &= \mathbb{P}(X_1 \in A, X_2 \in \mathbb{R}, \dots, X_d \in \mathbb{R}) \\ &= \int_{A \times \mathbb{R}^{d-1}} dx_1 \cdots dx_d \; f_X(x_1, \dots, x_d) \\ &= \int_A dx_1 \left[\int_{\mathbb{R}^{d-1}} dx_2 \cdots dx_d \; f_X(x_1, \dots, x_d) \right].\end{aligned}$$

La fonction définie par $f : x \mapsto \int \cdots \int_{\mathbb{R}^{d-1}} dx_2 \cdots dx_d \; f_X(x, x_2, \dots, x_d)$, est mesurable et intégrable (cf. théorème III.23). De plus on a $f \geq 0$, $\int f(x_1) dx_1 = 1$. Comme pour tout borélien A, $\mathbb{P}(X_1 \in A) = \int_A dx_1 \; f(x_1)$, on déduit de la définition, que X_1 est une v.a.c. dont la loi a pour densité f. □

Exercice III.1.
Montrer que si $X = (X_1, \dots, X_d)$ est une v.a.c. alors $(X_1, \dots, X_k)$ est une v.a.c. pour tout $k \leq d$. △

Remarque. À partir de la loi du couple (X, Y), on peut donc calculer la loi de X et la loi de Y. La réciproque est fausse comme le montre l'exercice suivant. ◊

Exercice III.2.
Soit deux couples (X_1, Y_1), dont la loi a pour densité $f_1(x, y) = (x+y)\mathbf{1}_{(x,y)\in[0,1]^2}$, et (X_2, Y_2), dont la loi a pour densité $f_2(x, y) = \left(x + \frac{1}{2}\right)\left(y + \frac{1}{2}\right)\mathbf{1}_{(x,y)\in[0,1]^2}$.
Montrer que les lois marginales sont égales ($\mathcal{L}(X_1) = \mathcal{L}(X_2)$ et $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$), alors que les lois des couples sont distinctes. △

III.3 Espérance

On admet qu'à partir des définitions II.18 et II.19 du paragraphe II.7, on peut démontrer la proposition suivante sur l'espérance des v.a.c.

Proposition III.9. *Soit X une v.a.c. à valeurs dans $\mathbb{R}^d$, et soit f la densité de sa loi. Soit φ une fonction mesurable réelle.*

1. *Si $\varphi(X)$ est p.s. positive alors on a :*

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\, dx \in [0+, \infty].$$

2. *Si $\varphi(X)$ est* ***intégrable*** *(i.e. $\int_{\mathbb{R}} |\varphi(x)|\, f(x)\, dx < \infty$), alors on a :*

$$\boxed{\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f(x)\, dx.}$$

Exemple. Si la loi de X est la loi uniforme sur $[0, 2\pi]$, alors $\mathbb{E}[X] = \pi$. ◊

Exemple. Soit X une v.a.c. à valeurs dans $\mathbb{R}^d$ et f la densité de sa loi. Si $A \in \mathcal{B}(\mathbb{R}^d)$, alors $\mathbf{1}_A(X)$ est intégrable, et on a :

$$\mathbb{E}[\mathbf{1}_A(X)] = \int_{\mathbb{R}^d} \mathbf{1}_A(x) f(x)\, dx = \int_A f(x)\, dx = \mathbb{P}(X \in A).$$

En particulier, on retrouve $\mathbb{E}[\mathbf{1}] = 1$. ◊

III.4 Lois usuelles

On donne des exemples importants de lois à densité.

Exemple III.10. Loi **uniforme** $U_{[a,b]}$ sur l'intervalle $[a, b]$:

$$f(x) = \frac{1}{b-a}\, \mathbf{1}_{[a,b]}(x).$$

Plus généralement, si $A \subset \mathbb{R}^d$ est mesurable de mesure de Lebesgue $\lambda(A) > 0$, alors la loi uniforme sur A a pour densité $f(x) = \dfrac{1}{\lambda(A)}\, \mathbf{1}_A(x)$. ◊

Exemple III.11. Loi **gaussienne** ou **normale** $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 :

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x-m)^2/2\sigma^2}, \quad x \in \mathbb{R}.}$$

L'intégrale de la fonction f est bien égale à 1 (voir l'exercice III.12). Cette loi a une importance capitale en calcul des probabilités et en statistique. Elle apparaît naturellement comme la loi d'erreurs de mesures par exemple (cf. le théorème central limite V.29). Elle permet de construire des modèles robustes pour lesquels on peut faire des calculs explicites. Le chapitre sur les vecteurs gaussiens est une

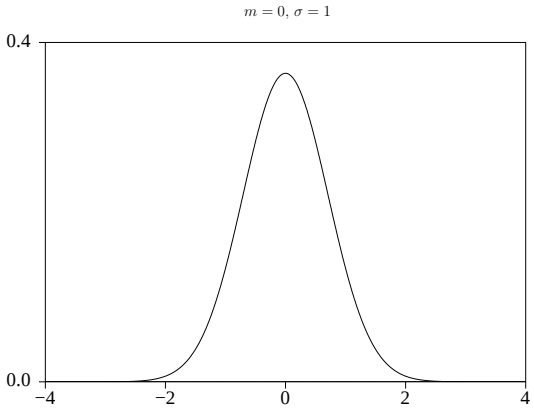


Figure III.2. Densité de la loi gaussienne $\mathcal{N}(m, \sigma^2)$.

généralisation au cas vectoriel de la loi gaussienne. Enfin les paramètres de cette loi ont une interprétation immédiate : m est la **moyenne** et σ^2 la **variance** comme l'indique le lemme suivant.

Lemme III.12. *Soit X de loi gaussienne $\mathcal{N}(m, \sigma^2)$. Alors X est de carré intégrable et on a $\mathbb{E}[X] = m$ et $\mathrm{Var}(X) = \sigma^2$.*

Démonstration. On remarque que comme $\mathrm{e}^{-y^2/2}/\sqrt{2\pi}$ est la densité de la loi $\mathcal{N}(0,1)$, on a :

$$\int_{\mathbb{R}} \mathrm{e}^{-y^2/2} \frac{dy}{\sqrt{2\pi}} = 1.$$

En posant $y = (x-m)/\sigma$, il vient :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-m)^2/2\sigma^2} \, dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (m + \sigma y) \mathrm{e}^{-y^2/2} \, dy = m.$$

On a également, à l'aide d'une intégration par partie :

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{\mathbb{R}} (x-m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-m)^2/2\sigma^2} \, dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 \mathrm{e}^{-y^2/2} \, dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[-y \mathrm{e}^{-y^2/2} \right]_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-y^2/2} \, dy \\
&= \sigma^2.
\end{aligned}$$

□

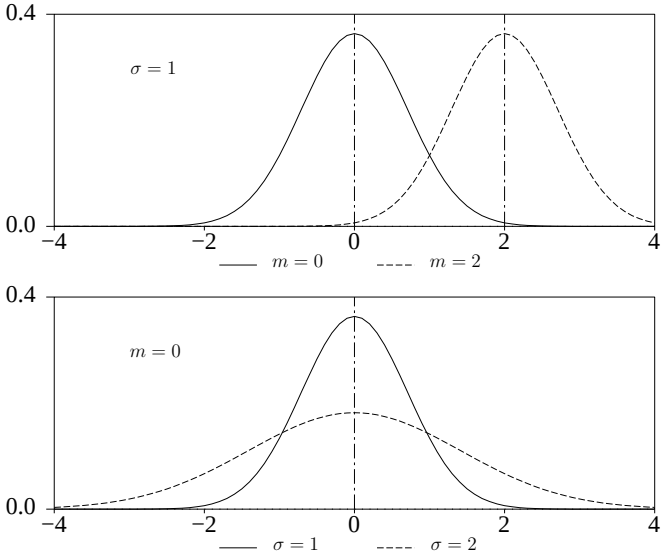


Figure III.3. Densités de lois gaussiennes $\mathcal{N}(m, \sigma^2)$.

Exercice III.3.
Montrer que si $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$, et si $\mathcal{L}(X) = \mathcal{N}(m, \sigma^2)$, alors la loi de $aX + b$ est la loi gaussienne $\mathcal{N}(am + b, a^2\sigma^2)$. △

Il découle de l'exercice précédent que la fonction de répartition de la loi $\mathcal{N}(m, \sigma^2)$ se déduit de celle de $\mathcal{N}(0, 1)$. Cette dernière est tabulée (voir le paragraphe XI.1). On retient que pour la loi $\mathcal{N}(0, 1)$, le quantile à 97,5% est $z \simeq 1.96$ et le quantile à 99.5% est $z \simeq 2.58$. En particulier, si $\mathcal{L}(X) = \mathcal{N}(m, \sigma^2)$, alors on a :

$$\mathbb{P}(X \in [m \pm 1.96\sigma]) \simeq 95\% \quad \text{et} \quad \mathbb{P}(X \in [m \pm 2.58\sigma]) \simeq 99\%.$$

Le graphe de la densité (cf. les figures III.2 et III.3) est une courbe en cloche très écrasée aux extrémités (décroissance en e^{-cx^2}). ◊

Exemple III.13. La loi **exponentielle** $\mathcal{E}(\lambda)$ de paramètre $\lambda \in]0, \infty[$:

$$\boxed{f(x) = \lambda\, \mathrm{e}^{-\lambda x}\, \mathbf{1}_{\{x \geq 0\}}.}$$

Voir la figure III.4 pour une représentation de ces densités. La loi exponentielle apparaît naturellement dans les modèles de files d'attente, de temps d'attente de pannes, de durée de vie de particules radioactives. Elle apparaît également comme la limite des lois géométriques (voir le chapitre V.3 sur les théorèmes de convergence en loi). À ce titre elle partage de nombreuses propriétés avec les lois géométriques. En particulier, la loi exponentielle n'a pas de "mémoire" (voir l'exercice III.26).

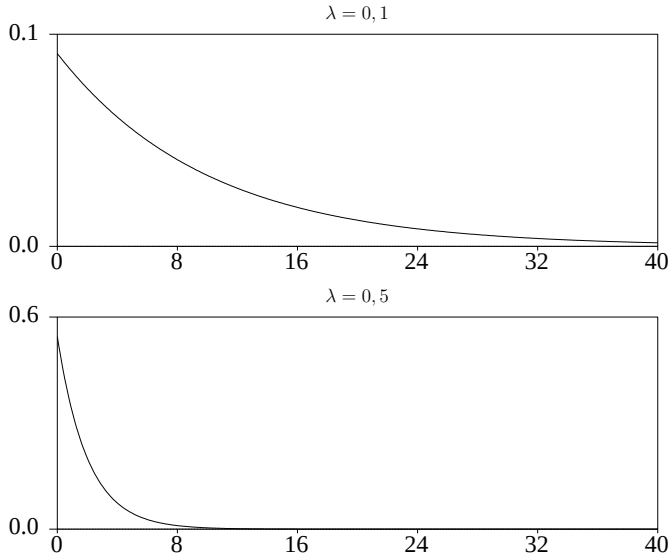


Figure III.4. Densité de la loi exponentielle de paramètre λ.

$\Diamond$

III.5 Autres lois

Les lois suivantes sont également d'un usage fréquent.

Exemple III.14. La loi de **Cauchy** de paramètre $a > 0$:

$$\boxed{f(x) = \frac{a}{\pi}\,\frac{1}{x^2 + a^2}, \quad x \in \mathbb{R}.}$$

Cette loi ne possède pas de moment d'ordre 1 : si X est de loi de Cauchy de paramètre a, alors X n'est pas intégrable. En effet l'intégrale $\int |x|\, f(x)dx$ est divergente en $+\infty$ et $-\infty$.

Le graphe de la densité, voir la figure III.5, est une courbe en cloche beaucoup moins écrasée aux extrémités (décroissance en x^{-2}) que le graphe de la densité gaussienne. En particulier cela implique que la probabilité qu'une variable gaussienne X de loi $\mathcal{N}(0,1)$ prenne de grandes valeurs est beaucoup plus faible que la probabilité qu'une variable aléatoire de Cauchy Y de paramètre 1. Ainsi on a :

$$\mathbb{P}(|X| > 6) \simeq 2.\,10^{-9} \quad \text{et} \quad \mathbb{P}(|Y| > 6) \simeq 0,1.$$

$\Diamond$

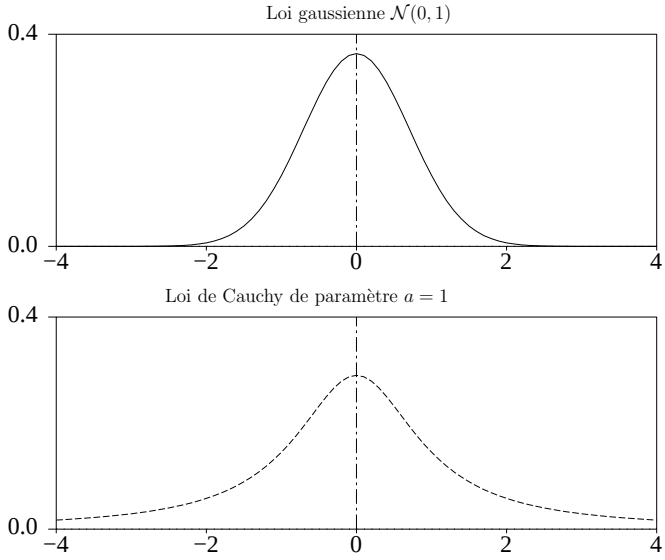


Figure III.5. Densité de la loi gaussienne et de la loi de Cauchy.

Exemple III.15. La loi **gamma** $\Gamma(\lambda, \alpha)$ de paramètre $\lambda > 0$ et $\alpha > 0$:

$$\boxed{f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} \, \mathrm{e}^{-\lambda x} \, \mathbf{1}_{\{x>0\}},}$$

où la fonction Gamma est définie par $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \, \mathrm{e}^{-x} \; dx$. Attention, la loi Γ de paramètre (λ, α) est parfois notée $\Gamma(\alpha, \lambda)$ ou $\Gamma(\alpha, 1/\lambda)$ suivant les ouvrages. Voir la figure III.6 pour une représentation de ces densités.

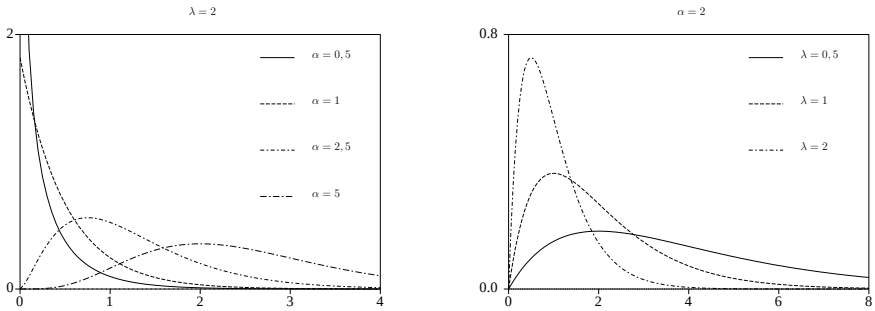


Figure III.6. Densités de lois gamma $\Gamma(\lambda, \alpha)$.

Pour $\alpha = 1$, on retrouve les lois exponentielles de paramètre λ. Enfin si $\lambda = 1/2$ et $\alpha = d/2$, on parle plutôt de loi du **khi 2** à d degrés de liberté notée $\chi^2(d)$. Cette loi apparaît naturellement dans les modèles gaussiens (paragraphe VI.2) et en statistique (paragraphe IX.9). Cette loi est tabulée (voir le paragraphe XI.3).

Concernant la fonction Γ, on rappelle que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ pour $\alpha > 0$, et $\Gamma(n) = (n-1)!$ si n est entier. On a $\Gamma(1) = 1$ et $\Gamma(1/2) = \sqrt{\pi}$. On donne également un équivalent de $n!$ et de $\Gamma(t)$ à l'infini (formule de Stirling) :

$$\lim_{n\to\infty} \frac{n!}{n^n\,\mathrm{e}^{-n}\,\sqrt{2\pi n}} = 1 \quad \text{et} \quad \lim_{t\to\infty} \frac{\Gamma(t)}{t^{t-\frac{1}{2}}\,\mathrm{e}^{-t}\,\sqrt{2\pi}} = 1.$$

◊

Exemple III.16. La loi **béta** $\beta(a,b)$ de paramètre $a > 0, b > 0$:

$$\boxed{f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbf{1}_{]0,1[}(x).}$$

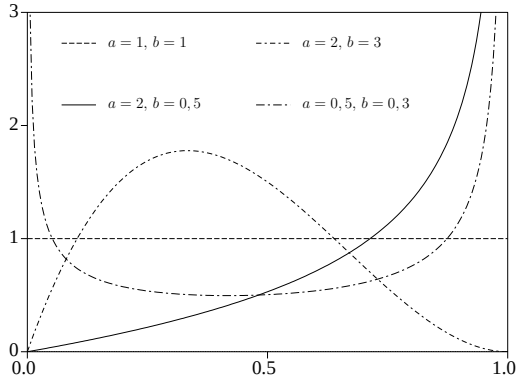


Figure III.7. Densités de lois béta $\beta(a,b)$.

Voir la figure III.7 pour une représentation de ces densités. La loi uniforme est la loi β de paramètre $(1,1)$. ◊

III.6 Indépendance

Dans le cas de v.a. continues, on peut donner une caractérisation de l'indépendance (définition II.32).

Proposition III.17. *Soit $X_1, \dots, X_n$, n variables aléatoires continues réelles ou vectorielles. Elles sont **indépendantes** si et seulement si le vecteur $(X_1, \dots, X_n)$ est une v.a.c. et sa densité s'exprime comme le produit des densités marginales : pour presque tout $(x_1, \dots, x_n)$,*

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

Démonstration. Pour simplifier, on suppose que les v.a. $X_1, \dots, X_n$ sont réelles. La démonstration se généralise sans difficulté à des v.a.c. vectorielles.

On suppose que $(X_1, \dots, X_n)$ est une v.a. continue et que $f_{X_1,\dots,X_n} = f_{X_1} \cdots f_{X_n}$ presque partout. Soit $f_1, \dots, f_n$ des fonctions réelles mesurables bornées. On a :

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] &= \int \cdots \int f_1(y_1) \cdots f_n(y_n)\ f_{X_1,\dots,X_n}(y_1, \dots, y_n)\ dy_1 \cdots dy_n \\
&= \int \cdots \int \prod_{i=1}^{n} f_i(y_i) f_{X_i}(y_i)\ dy_i \\
&= \prod_{i=1}^{n} \int f_i(y_i) f_{X_i}(y_i)\ dy_i \\
&= \prod_{i=1}^{n} \mathbb{E}\left[f_i(X_i)\right].
\end{aligned}$$

On a donc (II.9).

On démontre maintenant la réciproque. Soit $X_1, \dots, X_n$ des v.a.c. réelles indépendantes. On calcule la fonction de répartition de $X = (X_1, \dots, X_n)$. Pour $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, on a :

$$\begin{aligned}
F_X(x) &= \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \le x_i) \\
&= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(y_i)\ dy_i \\
&= \int_{\prod_{i=1}^{n}]-\infty, x_i]} \prod_{i=1}^{n} f_{X_i}(y_i)\ dy_1 \cdots dy_n,
\end{aligned}$$

où l'on a utilisé l'indépendance pour la deuxième égalité. La fonction $f(x) = \prod_{i=1}^{n} f_{X_i}(x_i)$ est positive d'intégrale 1 sur $\mathbb{R}^n$. Il s'agit donc d'une densité. Ainsi

la v.a. X a même fonction de répartition qu'une variable aléatoire dont la loi a pour densité f. Le théorème III.5 assure que X est une v.a.c. dont la loi a pour densité f. □

Exercice III.4.
On considère un gâteau circulaire avec une cerise sur le bord. On découpe le gâteau en deux parts en coupant suivant deux rayons choisis au hasard.

1. Avec quelle probabilité la part contenant la cerise est-elle plus petite que la part ne contenant pas la cerise ?
2. Quelle est la longueur angulaire moyenne de la part contenant la cerise ?

△

Correction III.4. On note Θ_1 et Θ_2 les angles formés par les deux rayons et le rayon qui passe par la cerise. L'énoncé du problème indique que Θ_1 et Θ_2 sont indépendants et suivent la loi uniforme sur $[0, 2\pi]$. La longueur angulaire de la part contenant la cerise est $2\pi - |\Theta_1 - \Theta_2|$.

1. La probabilité pour que la part contenant la cerise soit la plus petite est $\mathbb{P}(2\pi - |\Theta_1 - \Theta_2| < |\Theta_1 - \Theta_2|)$. Comme les angles sont indépendants, la loi du couple est la loi produit. On calcule :

$$\begin{aligned}\mathbb{P}(2\pi - |\Theta_1 - \Theta_2| < |\Theta_1 - \Theta_2|) &= \frac{1}{(2\pi)^2} \iint_{[0,2\pi]^2} \mathbf{1}_{\{|\theta_1-\theta_2|>\pi\}} d\theta_1 d\theta_2 \\ &= \frac{1}{(2\pi)^2}\, 2 \int_{[0,2\pi]} d\theta_1 \int_{[0,\theta_1]} \mathbf{1}_{\{\theta_1-\theta_2>\pi\}}\, d\theta_2 \\ &= \frac{1}{4}\,.\end{aligned}$$

   La probabilité pour que la part contenant la cerise soit la plus petite est $1/4$.
2. La longueur moyenne de la part contenant la cerise est égale à $2\pi - \mathbb{E}[|\Theta_1 - \Theta_2|]$. On calcule :

$$\begin{aligned}\mathbb{E}[|\Theta_1 - \Theta_2|] &= \frac{1}{(2\pi)^2} \iint_{[0,2\pi]^2} |\theta_1 - \theta_2|\ d\theta_1 d\theta_2 \\ &= \frac{1}{(2\pi)^2}\, 2 \int_{[0,2\pi]} d\theta_1 \int_{[0,\theta_1]} (\theta_1 - \theta_2)\, d\theta_2 \\ &= \frac{2\pi}{3}.\end{aligned}$$

   La longueur moyenne de la part contenant la cerise est donc $4\pi/3$.

La part qui contient la cerise est plus grande en moyenne et elle est également plus grande dans 75% des cas. Pour voir que ces résultats ne contredisent pas l'intuition il faut inverser les opérations. On découpe d'abord au hasard deux rayons dans le gâteau, puis on jette au hasard la cerise sur le bord. Celle-ci a intuitivement plus de chance de tomber sur la part la plus grosse ! Il reste à se convaincre que jeter la cerise sur le bord puis couper le gâteau au hasard, ou couper le gâteau au hasard puis jeter la cerise sur le bord donne bien le même résultat.

▲

III.7 Calcul de lois

Pour calculer une loi de variable aléatoire il existe plusieurs méthodes. On en verra essentiellement quatre :

- Les fonctions **génératrices** pour les v.a. discrètes à valeurs entières (paragraphe II.12).

- Les fonctions de **répartitions**, particulièrement adaptées pour calculer les lois de minimum ou maximum de variables aléatoires indépendantes (voir les exercices II.21 et III.16).

- La méthode de la **fonction muette** pour les v.a. continues que l'on détaille dans ce paragraphe.

- Les fonctions **caractéristiques** qui généralisent la notion de fonction génératrice. Cette méthode sera abordée au chapitre IV.

La méthode de la fonction muette repose sur la proposition suivante qui découle de la définition III.1 (prendre $g(x) = \mathbf{1}_A(x)$).

Proposition III.18. *Soit X une v.a. telle que pour toute fonction **bornée** mesurable g, on ait :*

$$\mathbb{E}[g(X)] = \int g(x)f(x)\,dx.$$

Alors X est une v.a. continue et la fonction f est la densité de sa loi.

Remarque. En fait on peut restreindre l'ensemble des fonctions tests g par exemple aux sous-ensembles suivants : fonctions indicatrices de pavé ; fonctions continues bornées ; fonctions exponentielles complexes $g(x) = \mathrm{e}^{iux}$, $u \in \mathbb{C}$ (cf. théorème

IV.5) ; fonctions exponentielles réelles $g(x) = \mathrm{e}^{-ux}$, $u \in \mathbb{R}^+$, pour des v.a. réelles positives (résultat admis). ♢

On utilise cette méthode dans l'exercice typique suivant.

Exercice III.5.
Soit Y une variable aléatoire de loi exponentielle $\lambda > 0$ et ε une variable aléatoire discrète indépendante de Y et telle que $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. Montrer que la variable aléatoire $Z = \varepsilon Y$ est à densité et la calculer. Cette loi est appelée loi exponentielle symétrique. △

Correction III.5. Soit g une fonction mesurable bornée. En utilisant $\varepsilon = \mathbf{1}_{\{\varepsilon=1\}} - \mathbf{1}_{\{\varepsilon=-1\}}$ p.s., puis l'indépendance, on a :

$$\begin{aligned}
\mathbb{E}[g(Z)] &= \mathbb{E}[g(Y)\mathbf{1}_{\{\varepsilon=1\}}] + \mathbb{E}[g(-Y)\mathbf{1}_{\{\varepsilon=-1\}}] \\
&= \mathbb{E}[g(Y)]\mathbb{P}(\varepsilon = 1) + \mathbb{E}[g(-Y)]\mathbb{P}(\varepsilon = -1) \\
&= \frac{1}{2}\int_0^{+\infty} \lambda\,\mathrm{e}^{-\lambda y}\, g(y)dy + \frac{1}{2}\int_0^{+\infty} \lambda\,\mathrm{e}^{-\lambda y}\, g(-y)dy \\
&= \frac{1}{2}\int_{\mathbb{R}} \lambda\,\mathrm{e}^{-\lambda\,|z|}\, g(z)\; dz.
\end{aligned}$$

Donc Z est une variable aléatoire continue de densité f définie par $f(z) = \frac{1}{2}\,\lambda\,\mathrm{e}^{-\lambda\,|z|}$, $z \in \mathbb{R}$. ▲

Dans la correction précédente, la fonction g est quelconque et ne joue aucun rôle direct, d'où le nom de méthode de la fonction muette. Dans de nombreux cas, la méthode de la fonction muette utilise un changement de variables vectorielles, voir par exemple la correction de l'exercice III.8. Ces changements de variables sont rappelés au paragraphe III.10.

Exercice III.6.
Soit X de loi $\mathcal{N}(0, \sigma^2)$. Calculer et reconnaître la loi de X^2. △

Exercice III.7.
Soit X de loi de Cauchy de paramètre a. Calculer la loi de $1/X$. △

Exercice III.8.
Soit X et Y deux variables aléatoires réelles indépendantes continues de densité f_X et f_Y. Montrer que la variable aléatoire $X + Y$ est continue de densité $f_X * f_Y$ où $*$ désigne le produit de convolution : pour tout $z \in \mathbb{R}$,

$$f_X * f_Y(z) = \int_{\mathbb{R}} dv\; f_X(z - v) f_Y(v).$$ △

Correction III.8. Comme les variables X et Y sont indépendantes, (X,Y) est une variable aléatoire continue de densité $(x,y) \mapsto f_X(x)f_Y(y)$. La fonction $(x,y) \mapsto (z,v) = (x+y,y)$ est un C^1 difféomorphisme de $\mathbb{R}^2$ dans lui-même. La matrice jacobienne de la transformation est $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. La valeur absolue du déterminant de la matrice jacobienne est constante, égale à 1. On en déduit que pour toute fonction g mesurable bornée définie sur $\mathbb{R}$, on a :

$$\begin{aligned}\mathbb{E}[g(X+Y)] &= \int g(x+y)\ f_X(x)f_Y(y)\ dxdy \\ &= \int g(z)\ f_X(z-v)f_Y(v)\ dzdv \\ &= \int g(z)\ f_X * f_Y(z)dz,\end{aligned}$$

où l'on a utilisé le théorème de Fubini pour la dernière égalité. Le résultat découle de la proposition III.18. ▲

III.8 Lois conditionnelles

On désire étendre la notion de loi conditionnelle vue pour les v.a.d. au paragraphe II.14 aux v.a.c. Soit (X,Y) un vecteur aléatoire continu. On ne peut pas calculer $\mathbb{P}(Y \in A | X = x)$ car $\mathbb{P}(X = x) = 0$. On a intuitivement si $f_{Y,X}$ et f_X représentent la densité de la loi de (Y,X) et de X :

$$\begin{aligned}\mathbb{P}(Y \in A | X \in [x, x+\Delta x]) &= \frac{\mathbb{P}(Y \in A, X \in [x, x+\Delta x])}{\mathbb{P}(X \in [x, x+\Delta x])} \\ &= \frac{\int \mathbf{1}_A(y)\mathbf{1}_{[x,x+\Delta x]}(z) f_{Y,X}(y,z)\ dydz}{\int \mathbf{1}_{[x,x+\Delta x]}(z) f_{Y,X}(y,z)\ dydz} \\ &= \frac{\int_x^{x+\Delta x} dz \int \mathbf{1}_A(y) f_{Y,X}(y,z)\ dy}{\int_x^{x+\Delta x} f_X(z)\ dz}.\end{aligned}$$

Si les densités sont continues et si $f_X(x) \neq 0$, en passant à la limite $\Delta x \to 0$, le membre de droite converge vers $\int_A [f_{Y,X}(y,x)/f_X(x)]\ dy$. On remarque que si $f_X(x) \neq 0$, la fonction de $y \mapsto f_{Y,X}(y,x)/f_X(x)$ est mesurable, positive et d'intégrale 1. Cela correspond donc à la densité d'une loi. Ceci motive la définition suivante.

Définition III.19. *Soit (X,Y) un vecteur aléatoire continu de densité $f_{(X,Y)}$. On note f_X la densité de la loi de X. Si $f_X(x) \neq 0$, alors la fonction de la variable y :*

$$\boxed{f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)}}$$

est une densité. Elle est appelée densité de la ***loi conditionnelle de*** Y ***sachant*** $X = x$.

Cette définition est valable pour des v.a.c. vectorielles. Et par convention on pose :

$$\boxed{\mathbb{P}(Y \in A|X = x) = \int_A f_{Y|X}(y|x)\, dy.}$$

On a la formule de Bayes :

$$\boxed{f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int f_{Y|X}(y|x) f_X(x)\, dx}.}$$

Soit φ une fonction mesurable. On suppose que $\varphi(X,Y)$ est intégrable. La fonction définie par :

$$\psi(x) = \int \varphi(x,y) f_{Y|X}(y|x)\, dy \text{ si } f_X(x) \neq 0 \quad \text{et } \psi(y) = 0 \text{ sinon,} \tag{III.3}$$

est mesurable. On remarque que $\psi(X)$ est intégrable. En effet, on a :

$$\begin{aligned}\int |\psi(x)|\, f_X(x)\, dx &\leq \int |\varphi(x,y)|\, f_{Y|X}(y|x) f_X(x)\, dxdy \\ &= \int |\varphi(x,y)|\, f_{X,Y}(x,y)\, dxdy < \infty.\end{aligned}$$

Définition III.20. *Soit (X,Y) une v.a.c. et φ une fonction mesurable telle que $\varphi(X,Y)$ est intégrable. La variable aléatoire $\psi(X)$, où ψ est définie par (III.3), est l'****espérance conditionnelle*** *de $\varphi(X,Y)$ sachant X. On la note $\mathbb{E}[\varphi(X,Y)|X]$. Par convention on note $\psi(x) = \mathbb{E}[\varphi(X,Y)|X = x]$.*

Exercice III.9.
On considère (X,Y) un vecteur aléatoire à valeurs dans $\mathbb{R}^2$, continu et de densité $f_{(X,Y)}(x,y) = \lambda x^{-1} \mathrm{e}^{-\lambda x} \mathbf{1}_{\{0<y<x\}}$. Déterminer la loi conditionnelle de Y sachant X. Soit φ une fonction réelle mesurable bornée. Calculer $\mathbb{E}[\varphi(Y)|X]$. △

Correction III.9. Par la formule des lois marginales, on calcule la densité de la loi de X : $f_X(x) = \int f_{X,Y}(x,y)\,dy = \lambda\,\mathrm{e}^{-\lambda x}\,\mathbf{1}_{\{x>0\}}$. On en déduit que, pour $x > 0$, $f_{Y|X}(y|x) = x^{-1}\mathbf{1}_{\{0<y<x\}}$. On reconnaît la densité de la loi uniforme sur $[0,x]$. On dit que conditionnellement à X, Y suit une loi uniforme sur $[0,X]$. Comme φ est bornée, $\varphi(Y)$ est intégrable et on a $\mathbb{E}[\varphi(Y)|X] = X^{-1}\int_0^X \varphi(y)\,dy$. ▲

Exercice III.10.
Soit X_1, X_2 des v.a.c. uniformes sur $[0,1]$ indépendantes. Calculer la loi de X_1 sachant $S = X_1 + X_2$. Pour cela on pourra calculer d'abord la loi du couple (X_1, S). Remarquer que la loi de X_1 sachant S est la loi uniforme sur l'intervalle $[S-1, S] \cap [0,1]$. △

Remarque III.21. On admet le résultat suivant qui est similaire à la proposition III.18. Soit X, Y deux v.a. telles que pour toute fonction **bornée** mesurable g, on ait :

$$\mathbb{E}[g(Y)|X] = \psi(X), \quad \text{où} \quad \psi(x) = \int g(y)h(x,y)\,dy.$$

Alors conditionnellement à $X = x$, Y est une v.a.c. de densité $y \mapsto h(x,y)$. ◊

Enfin l'espérance conditionnelle possède les propriétés de **linéarité**, **positivité** et **croissance** (cf. proposition II.49). La proposition II.48 et le lemme II.50 sont valables pour les v.a.c. (elles se démontrent facilement en utilisant le théorème de Fubini III.23). La proposition II.51 reste vraie avec (X,Z) v.a.c., $Y = \varphi(Y,Z)$ et $\mathbb{E}[Y|Z] = \mathbb{E}[\varphi(X,Z)|X]$ donné par la définition III.20.

III.9 Simulation

Afin d'étudier un modèle aléatoire, il peut être intéressant d'en faire des simulations. On verra également, grâce à la loi forte des grands nombres (voir le théorème V.24), que pour calculer l'espérance de $\varphi(X)$, on peut utiliser la méthode de Monte Carlo (voir l'exemple V.25) : i.e. approcher $\mathbb{E}[\varphi(X)]$ par $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\varphi(x_k)$, où x_k est la réalisation d'une v.a. X_k, et les v.a. $X_1, \ldots, X_k$ sont indépendantes et de même loi que X. En particulier, il est important d'obtenir des réalisations de suites de v.a. indépendantes.

Les générateurs de nombres pseudo-aléatoires fournissent des suites $(u_n, n \geq 1)$ que l'on admet être la réalisation d'une suite de v.a. indépendantes de loi uniforme

sur $[0,1]$. La construction des suites pseudo-aléatoires dépasse le cadre de ce cours. Une méthode élémentaire (mais typique) est de considérer la suite $(y_n/c, n \geq 1)$ où $y_0 \leq c$ est un entier, et pour $n \geq 0$:

$$y_{n+1} = ay_n + b \mod (c),$$

pour des entiers a, b et c bien choisis. La suite $(y_n/c, n \geq 1)$ est périodique, mais les premiers termes "ressemblent" à la réalisation de v.a. indépendantes de loi uniforme sur $[0,1]$. Les suites pseudo-aléatoires sont périodiques, mais en pratique les périodes dépassent très largement le nombre de termes utilisés de la suite.

La problématique de la simulation [1] d'une v.a. de loi donnée, est d'exhiber une v.a. de même loi qui s'exprime le plus simplement possible comme une fonction d'une ou plusieurs variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

Les exemples qui suivent permettent de donner des représentations particulières pour les lois usuelles.

Soit $(U_n, n \geq 1)$ une suite de v.a. indépendantes de loi uniforme sur $[0,1]$.
- Loi de Bernoulli de paramètre p : $\mathbf{1}_{\{U_1 \leq p\}}$.
- Loi binomiale de paramètre (n,p) : $\sum_{k=1}^{n} \mathbf{1}_{\{U_k \leq p\}}$.
- Loi géométrique de paramètre p : $\lceil \log(U_1)/\log(1-p) \rceil$ où $\lceil x \rceil \in \mathbb{Z}$ est le plus petit entier relatif plus grand que x (se déduit de la proposition III.22 et de l'exercice III.29).
- Loi uniforme sur $[a,b]$: $(b-a)U + a$.
- Loi de Poisson : voir l'exercice III.32.
- Loi exponentielle de paramètre λ : $-\log(U_1)/\lambda$ (conséquence directe de la proposition III.22).
- Loi gaussienne : voir les exercices III.33 et III.34.
- Loi de Cauchy : $\tan(2\pi U_1)$ (voir les exercices III.20 et III.33).

La méthode d'inversion de la fonction de répartition décrite dans la proposition suivante est une méthode générale de simulation pour les variables aléatoires réelles.

Proposition III.22 (Méthode d'inversion de la fonction de répartition)**.** *Soit X une variable aléatoire réelle de fonction de répartition F et F^{-1} son inverse généralisé, voir la définition III.6. Soit U une variable aléatoire de loi uniforme sur $[0,1]$. Alors la variable aléatoire $F^{-1}(U)$ a même loi que X.*

1. Voir : L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, 1986.

Démonstration. La manipulation des inverses généralisés est technique. On donne seulement une démonstration dans le cas particulier ou F est une bijection. Dans ce cas l'inverse généralisé et l'inverse coïncident.

Soit U de loi uniforme sur $[0,1]$. Pour $x \in \mathbb{R}$, on a $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$ et donc :

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Les v.a. $F^{-1}(U)$ et X ont donc même fonction de répartition. Elles ont donc même loi d'après le théorème III.5. □

La méthode du rejet présentée dans l'exercice III.35, voir aussi l'exercice III.34, est également très utilisée.

III.10 Rappels sur l'intégration

Dans ce chapitre on rappelle les théorèmes de convergence pour l'intégration par rapport à la mesure de Lebesgue, ainsi que les théorèmes de changement de variable.

Théorème III.23 (Fubini)**.**

1. *Soit $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}^+$ mesurable par rapport à la tribu borélienne. Alors pour tout $i \in \{1, \ldots, n\}$, la fonction f_i réelle définie sur $\mathbb{R}^{n-1}$ par :*

$$f_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \int_{\mathbb{R}} f(x_1, \ldots, x_n) dx_i$$

est mesurable. De plus, pour toute permutation σ de $\{1, \ldots, n\}$, les intégrales :

$$\int_{\mathbb{R}} dx_{\sigma(1)} \left[\cdots \int_{\mathbb{R}} dx_{\sigma(n)} f(x_1, \ldots, x_n) \right] \cdots \right] \tag{III.4}$$

sont toutes égales. (Elles peuvent prendre la valeur $+\infty$.)

2. *Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$ mesurable par rapport à la tribu borélienne et telle que l'intégrale (III.4) avec f remplacé par $|f|$ soit finie. Alors pour toute permutation σ de $\{1, \ldots, n\}$, les intégrales (III.4) sont bien définies, finies et toutes égales.*

Par convention, sous les hypothèses du théorème III.23, les valeurs communes de (III.4) sont notées $\int_{\mathbb{R}^n} f(x_1, \ldots, x_n)\ dx_1 \cdots dx_n$, et de manière plus concise $\int_{\mathbb{R}^n} f(x)\ dx$. On dit que f est **intégrable** si $\int_{\mathbb{R}^n} |f(x)|\ dx$ est fini. En particulier le théorème III.23 assure que $\int_{\mathbb{R}^n} f(x)\ dx$ est bien défini si f est positive ou intégrable.

Exercice.
Montrer que :

$$\int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2} dy \right) dx = \frac{\pi}{4}.$$

En déduire que la fonction $f(x,y) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$ n'est pas intégrable sur $[0,1]\times[0,1]$.
(On pourra d'abord calculer $\dfrac{\partial}{\partial y}\dfrac{y}{x^2+y^2}$.) △

On rappelle la définition de la fonction indicatrice. Soit A un sous-ensemble borélien de $\mathbb{R}^d$. La fonction $\mathbf{1}_A$ est définie par :

$$\begin{aligned} \mathbf{1}_A : \mathbb{R}^d &\to \{0,1\} \\ x &\mapsto \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

Il est d'usage de simplifier le plus possible les notations. Si par exemple $A = \{(x,y); x > y\} \in \mathbb{R}^2$, alors on notera $\mathbf{1}_{\{x>y\}}$ pour $\mathbf{1}_A(x,y)$.

Exercice.
Vérifier que $\int_{[0,1]^2} \mathbf{1}_{\{x>y\}}\, dxdy = 1/2$. △

Définition III.24. *On dit qu'une propriété définie sur $\mathbb{R}^d$ est vraie presque partout (noté p.p.) si l'ensemble A sur lequel elle n'est pas vérifiée est de mesure nulle (i.e. $\int_{\mathbb{R}^d} \mathbf{1}_A(x)\, dx = 0$).*

Théorème III.25 (Convergence dominée)**.** *Soit g une fonction mesurable positive définie sur $\mathbb{R}^d$ et une suite $(f_n, n \in \mathbb{N}^*)$ de fonctions réelles mesurables définies sur $\mathbb{R}^d$ telles que pour presque tout $x \in \mathbb{R}^d$, $|f_n(x)| \leq g(x)$ (les fonctions $|f_n|$ sont dominées par la fonction g) et pour tout $n \in \mathbb{N}$:*

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \textit{presque partout.}$$

*Si g est **intégrable**, alors les fonctions f et f_n sont intégrables et on a :*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f_n(x)dx = \int_{\mathbb{R}^d} f(x)dx.$$

Exercice.
Soit $f_n(x) = \mathbf{1}_{[n,n+1]}(x)$. Vérifier que : $\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)dx \neq \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)dx$. △

Théorème III.26 (Convergence monotone)**.** *Soit $(f_n, n \in \mathbb{N}^*)$ une suite de fonctions réelles mesurables* ***positives*** *définies sur $\mathbb{R}^d$ telles que pour tout $n \leq m \in \mathbb{N}^*$, on ait $f_n \leq f_m$ presque partout. Alors on a :*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f_n(x)dx = \int_{\mathbb{R}^d} \lim_{n\to\infty} f_n(x)dx.$$

Les limites peuvent prendre la valeur $+\infty$.

Exercice.
Soit $f_n(x) = -n^{-1}|x|$. Vérifier que : $\lim_{n\to\infty} \int_{\mathbb{R}^d} f_n(x)dx \neq \int_{\mathbb{R}^d} \lim_{n\to\infty} f_n(x)dx$. △

Théorème III.27 (Lemme de Fatou)**.** *Soit $(f_n, n \in \mathbb{N}^*)$ une suite de fonctions réelles* ***positives*** *mesurables définies sur $\mathbb{R}^d$. On a :*

$$\liminf_{n\to\infty} \int_{\mathbb{R}^d} f_n(x)dx \geq \int_{\mathbb{R}^d} \liminf_{n\to\infty} f_n(x)dx.$$

Exercice.
Vérifier le lemme de Fatou pour la suite de fonctions $f_n(x) = \mathbf{1}_{[n,n+1]}(x)$. △

Remarque III.28. On rappelle enfin que si A est de mesure de Lebesgue nulle, alors pour toute fonction mesurable $\int_A f(x)\,dx = 0$. ◊

On aborde maintenant les formules de changement de variable. On rappelle dans un premier temps la formule de changement de variable en dimension $d = 1$, puis son extension en dimension $d \geq 1$.

Théorème III.29. *Soit f une fonction d'un ouvert $I \subset \mathbb{R}$ dans $\mathbb{R}$, mesurable intégrable. Soit φ une bijection de I dans un ouvert $J \subset \mathbb{R}$ de classe C^1 et telle que φ^{-1} est de classe C^1. On a :*

$$\int_I f(x)dx = \int_J \frac{f(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}dy = \int_J f(\varphi^{-1}(y)) \left|\left(\varphi^{-1}\right)'(y)\right| dy.$$

Si on pose $g = \varphi^{-1}$, la dernière égalité s'écrit :

$$\int_I f(x)dx = \int_J f(g(y)) \left|g'(y)\right| dy.$$

Remarquer que les valeurs absolues $|g'(y)|$ permettent d'oublier l'ordre des bornes de l'ensemble d'intégration.

Exercice.
Calculer $\displaystyle\int_{\mathbb{R}} \mathrm{e}^{-\left|v+\frac{1}{v}\right|}\, dv$. (On pourra poser $2h = v + \frac{1}{v}$). △

Soit O et O' deux ouverts de $\mathbb{R}^d$. On dit que $\varphi = (\varphi_1, \ldots, \varphi_d)$ est un C^1 difféomorphisme de O dans O' si φ est une bijection de O dans O' de classe C^1, et si sa réciproque est également de classe C^1. On définit la matrice jacobienne de φ, $\mathrm{Jac}[\varphi](x)$, comme étant la dérivée $\dfrac{\partial\varphi}{\partial x}(x)$: c'est la matrice $(a_{i,j}(x))_{1\leq i,j,\leq d}$ où :

$$a_{i,j}(x) = \frac{\partial\varphi_i}{\partial x_j}(x), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Le jacobien de φ est le déterminant de la matrice jacobienne. On note $|\mathrm{Jac}[\varphi](x)|$ la valeur absolue du déterminant de $\mathrm{Jac}[\varphi](x)$. En dimension 1, on a $\mathrm{Jac}[\varphi](x) = \varphi'(x)$.

Théorème III.30. *Soit f une fonction d'un ouvert $O \subset \mathbb{R}^d$ dans $\mathbb{R}^l$, mesurable intégrable. Soit φ un C^1 difféomorphisme de O dans un ouvert O' de $\mathbb{R}^d$. On a :*

$$\int_O f(x)dx = \int_{O'} \frac{f(\varphi^{-1}(y))}{|\mathrm{Jac}[\varphi](\varphi^{-1}(y))|} dy = \int_{O'} f(\varphi^{-1}(y)) \left|\mathrm{Jac}[\varphi^{-1}](y)\right| dy.$$

Si on pose $g = \varphi^{-1}$, la dernière égalité s'écrit :

$$\int_O f(x)dx = \int_{O'} f(g(y)) \left|\mathrm{Jac}[g](y)\right| dy.$$

Exercice III.11.
Montrer que si f est intégrable sur $\mathbb{R}^2$, alors on a :

$$\iint_{x>0,y>0} f(x+y, x-y)\, dxdy = \frac{1}{2} \iint f(w,v) \mathbf{1}_{\{w>0\}} \mathbf{1}_{\{-w<v<w\}}\, dwdv.$$

△

Correction III.11. On pose $O =]0,\infty[\times]0,\infty[$, $\varphi(x,y) = (v,w) = (x-y, x+y)$. On vérifie facilement que φ est un C^1 difféomorphisme de O dans O', où $O' = \{(v,w); w > 0 \text{ et } -w < v < w\}$. Le jacobien de φ est :

$$\mathrm{Jac}[\varphi](x,y) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

On a $|\mathrm{Jac}[\varphi](x,y)| = 2$. Il vient :

$$\iint_{x>0,y>0} f(x+y, x-y)\, dxdy = \frac{1}{2} \iint_{O'} f(w,v)\, dwdv.$$

▲

Exercice III.12.
Montrer que $A = \int_{\mathbb{R}} \mathrm{e}^{-x^2/2}\, dx = (2\pi)^{-1/2}$. On calculera A^2 en utilisant un changement de variable en coordonnées polaires. On fera particulièrement attention pour déterminer les ensembles O et O'. On pourra se servir de la remarque III.28. △

III.11 Résumé

Soit X une variable aléatoire continue (v.a.c.) réelle ou vectorielle de densité $f(x)$.

- Pour tout borélien A, on a $\mathbb{P}(X \in A) = \int_A f(x)\, dx$.
- Si Y est une v.a. réelle quelconque, la **fonction de répartition** de Y est $F(y) = \mathbb{P}(Y \leq y)$. Si X est réelle, la fonction de répartition de X est $F(y) = \int_{-\infty}^{y} f(x)\, dx$.
- On dit que $\varphi(X)$ est **intégrable** si $\int |\varphi(x)|\, f(x)\ dx < \infty$. Si $\varphi(X)$ est intégrable, alors l'espérance de $\varphi(X)$ est :

$$\mathbb{E}[\varphi(X)] = \int \varphi(x) f(x)\, dx.$$

Soit (X,Y) une v.a.c.

- Les 3 propriétés suivantes sont équivalentes :
  1. X et Y sont **indépendantes**.
  2. $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$ presque partout (la loi du couple (X,Y) est la **loi produit**).
  3. $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$ pour toutes fonctions g, h mesurables bornées.

- Les équivalences ci-dessus s'étendent à un nombre fini de variables.
- **Formule des lois marginales**. On peut calculer la loi de X à partir de la loi du couple (X,Y) : $f_X(x) = \int f_{X,Y}(x,y)\,dy$. On ne peut pas en général retrouver la loi du couple à partir des lois marginales.
- La densité de la **loi conditionnelle** de Y sachant $X = x$ est, pour x tel que $f_X(x) \neq 0$:
$$f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)}.$$
- Si $\varphi(X,Y)$ est intégrable, l'**espérance conditionnelle** de $\varphi(X,Y)$ sachant X est $\psi(X)$ où $\psi(x) = \int \varphi(x,y) f_{Y|X}(y|x)\,dy$. Et on a $\mathbb{E}\big[\mathbb{E}[\varphi(X,Y)|X]\big] = \mathbb{E}[\varphi(X,Y)]$.
- Si h est une fonction mesurable bornée, alors on a $\mathbb{E}[h(X)\varphi(X,Y)|X] = h(X)\mathbb{E}[\varphi(X,Y)|X]$.
- Si Y est indépendant de X, la loi conditionnelle de Y sachant X est la loi de Y. Si de plus $g(Y)$ est intégrable, alors $\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)]$.
- L'espérance conditionnelle comme l'espérance possède les propriétés de linéarité, positivité et croissance.
- Si $g(Y)$ est une v.a. réelle de carré intégrable, alors $\mathbb{E}[g(Y)|X]$ est de carré intégrable et de plus on a $\mathbb{E}[g(Y)|X]^2 \leq \mathbb{E}[g(Y)^2|X]$.
- Récapitulatif des lois usuelles :
  - Loi **uniforme** $U_{[a,b]}$ sur l'intervalle $[a,b]$:
$$f(x) = \frac{1}{b-a}\,\mathbf{1}_{[a,b]}(x).$$
  - Loi **gaussienne** ou **normale** $\mathcal{N}(m,\sigma^2)$ de moyenne $m \in \mathbb{R}$ et de variance $\sigma^2 > 0$:
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-(x-m)^2/2\sigma^2}.$$
  - Loi **exponentielle** $\mathcal{E}(\lambda)$, $\lambda \in]0,\infty[$:
$$f(x) = \lambda\,\mathrm{e}^{-\lambda x}\,\mathbf{1}_{\{x>0\}}.$$
  - Loi de **Cauchy** de paramètre $a > 0$:
$$f(x) = \frac{a}{\pi}\,\frac{1}{x^2+a^2}.$$
  - Loi **gamma** $\Gamma(\lambda,\alpha)$, $\lambda > 0$ et $\alpha > 0$:
$$f(x) = \frac{1}{\Gamma(\alpha)}\,\lambda^\alpha x^{\alpha-1}\,\mathrm{e}^{-\lambda x}\,\mathbf{1}_{\{x>0\}}.$$

– Loi du **khi** 2 à d degrés de liberté $\chi^2(d)$, $d \in \mathbb{N}^*$: c'est la loi gamma $\Gamma(1/2, d/2)$.
– Loi **béta** $\beta(a,b)$, $a > 0, b > 0$:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbf{1}_{]0,1[}(x).$$

Loi	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
$U_{[a,b]}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{N}(m,\sigma^2)$	m	σ^2
$\mathcal{E}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Cauchy (a)	non défini	non défini
$\Gamma(\lambda,\alpha)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\chi^2(d)$	d	$2d$
$\beta(a,b)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$

Pour déterminer la loi d'une variable aléatoire, on utilisera le plus souvent l'une des méthodes suivantes :
– La **fonction génératrice** si la v.a. est à valeurs entières.
– La **fonction de répartition**, surtout si l'on calcule la loi d'un minimum ou d'un maximum de v.a. réelles indépendantes.
– La **fonction muette** de manière générale.
– La **fonction caractéristique** (cf le chapitre IV).

III.12 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice III.1 p. 84, *Exercice* III.2 p. 84, *Exercice* III.3 p. 87, *Exercice* III.4 p. 92, *Exercice* III.5 p. 94, *Exercice* III.6 p. 94, *Exercice* III.7 p. 94, *Exercice* III.8 p. 94, *Exercice* III.9 p. 96, *Exercice* III.10 p. 97, *Exercice* III.11 p. 102, *Exercice* III.12 p. 103.

Exercice III.13.
Soit X une variable aléatoire continue de densité $cx(1-x)\mathbf{1}_{[0,1]}(x)$. Déterminer c, puis calculer $\mathbb{E}[X]$, $\mathrm{Var}(X)$ et $\mathbb{E}[1/X]$. △

Exercice III.14.
On considère un bâton sur lequel on trace au hasard deux marques. On découpe le bâton suivant les deux marques. Quelle est la probabilité pour que l'on puisse faire un triangle avec les trois morceaux ainsi obtenus ? △

Exercice III.15.
Votre ami choisit deux nombres positifs sans vous faire part de la manière dont il les choisit. Après avoir lancé une pièce équilibrée, il vous donne le plus petit s'il a obtenu face ou le plus grand s'il a obtenu pile. Vous devez parier s'il vous a donné le plus petit ou le plus grand. L'objectif est de maximiser la probabilité de gagner.

1. Vous lancez une pièce équilibrée ou non. Si vous obtenez face, vous pariez qu'il vous a donné le plus petit, sinon vous pariez qu'il vous a donné le plus grand. Quelle est la probabilité de gagner votre pari ?
2. Soit Z une variable aléatoire positive continue ayant pour support $\mathbb{R}^+$ (i.e. pour tout ouvert O de $\mathbb{R}^+$ non vide, on a $\mathbb{P}(Z \in O) > 0$). Si le nombre donné par votre ami est plus petit que Z, alors vous pariez qu'il vous a donné le plus petit, sinon vous pariez qu'il vous a donné le plus grand. Quelle est la probabilité de gagner votre pari ?
3. On suppose que les deux nombres de votre ami ont été obtenus par simulation suivant une loi (continue de densité strictement positive sur $]0,\infty[$) donnée et connue de vous. Déterminer votre stratégie optimale (i.e. la loi de Z que l'on ne suppose plus continue). Quelle est alors la probabilité de gagner votre pari ?

△

Exercice III.16.
Soit $Y_1, \cdots, Y_n$ des variables aléatoires indépendantes exponentielles de paramètre 1. Calculer, en utilisant la fonction de répartition, la loi de $V = \max_{i=1,\cdots,n} Y_i$. Calculer et reconnaître la loi de $W = \min_{i=1,\cdots,n} Y_i$. △

Exercice III.17.
Soit X, Y des variables aléatoires indépendantes de loi $\Gamma(\lambda, t)$ et $\Gamma(\lambda, s)$.

1. Calculer et reconnaître la loi de $\left(\frac{X}{X+Y}, X+Y\right)$.
2. On suppose $s > t$. Soit Z une variable indépendante de Y et de loi béta de paramètre $(t, s-t)$. Montrer que ZY a même loi que X.

△

Exercice III.18.
Généralisation de l'exercice III.17. Soit $(X_i, i \in \mathbb{N}^*)$ une suite de v.a. indépendantes telles que $\mathcal{L}(X_i) = \Gamma(\lambda, \alpha_i)$. Déterminer par récurrence et sans calcul la loi du vecteur :

$$\left(X_1 + \cdots + X_n, \frac{X_1}{X_1 + X_2}, \ldots, \frac{X_1 + \cdots + X_{n-1}}{X_1 + \cdots + X_n}\right).$$

△

Exercice III.19.
On modélise les temps d'attente aux caisses d'un supermarché par des variables aléatoires continues $(X_n, n \in \mathbb{N})$ positives, indépendantes et de même loi. Votre temps d'attente est représenté par X_0, et le temps d'attente de la personne arrivée en même temps que vous à la caisse n est X_n. Calculer la loi de $N = \inf\{n \geq 1; X_n > X_0\}$. Donner son espérance. Commentaire. △

Exercice III.20.
Soit X et Y des variables indépendantes de loi $\mathcal{N}(0,1)$.

1. Calculer la loi de X/Y.
2. En déduire la loi de $1/Z$ où Z est une variable aléatoire de Cauchy.

△

Exercice III.21.
Soit $X_1, \ldots, X_n$ des variables indépendantes de loi $\mathcal{N}(m, \sigma)$. Calculer la loi de $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ ainsi que son espérance et sa variance. △

Exercice III.22.
Soit X une v.a. positive de densité f. Montrer que :

$$\mathbb{E}[X^r] = \int_0^\infty r x^{r-1} \mathbb{P}(X > x) dx \quad \text{si} \quad r > 0.$$

Donner une formule analogue pour $r < 0$. △

Exercice III.23.
Si Z est une variable aléatoire de Cauchy pour quelles valeurs de $\alpha \in \mathbb{R}$ est-ce que $|Z|^\alpha$ est intégrable ? △

Exercice III.24.
Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ de densité f. Soit O un ouvert de $\mathbb{R}^d$ tel que $\mathbb{P}(X \in O) = 1$. Soit $\varphi \in C^1(O, O')$ telle que $\varphi^{-1} \in C^1(O', O)$, où O' est un ouvert de $\mathbb{R}^d$. Calculer la loi de $Y = \varphi(X)$. △

Exercice III.25.
Soit X de loi gaussienne centrée réduite $\mathcal{N}(0, 1)$.

1. Calculer $\mathbb{E}[X^n]$, pour $n \in \mathbb{N}$. (On pourra établir une formule de récurrence.)
2. Calculer $\mathbb{E}[e^{\lambda X}]$ pour $\lambda \in \mathbb{R}$. Indiquer comment ce calcul permet de retrouver formellement le résultat de la question précédente.

△

Exercice III.26.
On modélise la durée de fonctionnement entre deux de pannes d'une machine par des variables aléatoires sans mémoire :

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s) \quad t \geq 0, s \geq 0.$$

1. Montrer que les lois exponentielles sont sans mémoire.
2. Déterminer les variables aléatoires positives dont la loi admet une densité, qui sont sans mémoire. On montrera que la fonction $\bar{F}(t) = \mathbb{P}(X > t)$ satisfait une équation différentielle que l'on justifiera et que l'on résoudra.

△

Exercice III.27.
Soit X et Y deux variables aléatoires indépendantes de loi exponentielle de paramètre λ et μ.

1. Calculer la loi de $(\inf(X, Y), Z)$ où $Z = \max(X, Y) - \inf(X, Y)$. Quelle est la loi de Z si $\lambda = \mu$?
2. Montrer que les variables aléatoires $\inf(X, Y)$ et $\mathbf{1}_{\{X<Y\}}$ sont indépendantes. Donner la loi de $\mathbf{1}_{\{X<Y\}}$.

△

Exercice III.28.
On considère une suite $T_1, \ldots, T_n$ de variables aléatoires indépendantes de loi exponentielle de paramètre respectif $\alpha_1, \ldots, \alpha_n$.

1. Montrer que $\mathbb{P}(\text{Il existe } i \neq j \text{ tel que } T_i = T_j) = 0$.

On en déduit que l'on peut définir p.s. la variable aléatoire discrète M par :

$$M = i \Longleftrightarrow T_i < T_j \quad \text{pour tout } j \neq i.$$

On pose alors :

$$S = \min_{1 \leq k \leq n} T_k = T_M, \quad R_k = T_k - S \quad \text{pour } 1 \leq k \leq n.$$

2. Soit $i \in \{1, \ldots, n\}$. On considère des fonctions bornées mesurables f et g_j, pour $j \in \{1, \ldots, n\}$. Calculer $\mathbb{E}[\mathbf{1}_{\{M=i\}} f(S) \prod_{j \neq i} g_j(R_j)]$.
3. En déduire la loi de M et la loi de S.
4. Vérifier que M et S sont indépendants.
5. Montrer que conditionnellement à $\{M = i\}$, les variables aléatoires S, $R_1, \ldots,$ R_n, sont indépendantes. Déterminer la loi de R_j conditionnellement à $\{M = i\}$ pour $j \neq i$.

△

Exercice III.29.
Les lois exponentielles apparaissent comme des lois limites pour des lois géométriques changées d'échelle. On peut mettre en évidence d'autres propriétés.

1. Montrer que si T est une variable aléatoire exponentielle de paramètre λ, alors $[Tm] + 1$, où $[x]$ représente la partie entière de x, et $m > 0$, est une variable aléatoire géométrique dont on déterminera le coefficient.
2. Soit T une variable aléatoire positive telle que $[T2^n]+1$ est une v.a. géométrique pour tout $n \in \mathbb{N}^*$. On note p_n son paramètre. Établir une relation de récurrence entre $q_n = 1 - p_n$ et $q_{n+1} = 1 - p_{n+1}$. Montrer que :

   $$\{[T2^n] + 1 \geq [2^n x] + 2\} \subset \{T \geq x\} \subset \{[T2^n] + 1 \geq [2^n x]\}.$$

   En conclure que T suit une loi exponentielle. On déterminera son paramètre en fonction de q_0.
3. Soit T une variable aléatoire positive. On suppose qu'il existe une suite $(m_n, n \geq 1)$ croissante avec $m_0 > 0$ et $\lim_{n\to\infty} m_n = \infty$, et que pour tout $n \geq 1$, $[Tm_n] + 1$ est une variable aléatoire géométrique. Montrer que T suit une loi exponentielle. On exprimera son paramètre à l'aide de celui de $[Tm_0]$. (On montrera d'abord que $\mathbb{P}(T > x) = \lim_{n\to\infty} \mathbb{P}([Tm_n] > [m_n x])$ pour $x \geq 0$.)

△

Exercice III.30.
Soit T une variable aléatoire exponentielle de paramètre $\lambda > 0$. On suppose que la loi de S conditionnellement à T est la loi de Poisson de paramètre θT, avec $\theta > 0$. Déterminer et reconnaître la loi de $S + 1$. △

Exercice III.31.
Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes. On suppose que la loi de X_n est la loi exponentielle de paramètre $\lambda_n > 0$. Montrer que les trois assertions suivantes sont équivalentes.

(i) $\mathbb{P}(\sum_{n\geq 1} T_n = \infty) > 0$.
(ii) $\sum_{n\geq 1} {\lambda_n}^{-1} = \infty$.
(iii) $\mathbb{P}(\sum_{n\geq 1} T_n = \infty) = 1$.

Pour montrer que (ii) implique (iii), on considérera $\mathbb{E}[\exp(-\sum_{n\geq 1} X_n)]$. △

Exercice III.32.
Soit $(U_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Soit $\theta > 0$.

1. Donner la loi de $X_k = -\log(U_k)/\theta$.
2. Donner la loi de $\sum_{k=1}^n X_k$.
3. Calculer la loi de N défini par $N = \inf\left\{n; \prod_{k=1}^{n+1} U_k < \mathrm{e}^{-\theta}\right\}$.
4. En déduire une méthode pour simuler des variables aléatoires de Poisson.

△

Exercice III.33.
Soit X, Y des variables indépendantes de loi $\mathcal{N}(0,1)$.

1. On pose $R = \sqrt{X^2+Y^2}$ et $\Theta \in [0, 2\pi[$, définis par $R\cos\Theta = X$ et $R\sin\Theta = Y$. Calculer la loi de (R, Θ). En déduire que R et Θ sont indépendants.
2. Reconnaître les lois de R^2, $\mathrm{e}^{-R^2/2}$ et de $\tan\Theta$.
3. Soit Z, S des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Déduire des questions précédentes la loi du couple (X', Y') défini par :

$$X' = \cos(2\pi Z)\sqrt{2\,|\log S|} \quad \text{et} \quad Y' = \sin(2\pi Z)\sqrt{2\,|\log S|}.$$

Il s'agit de la transformation de Box-Muller. △

Exercice III.34.
Soit $((U_i, V_i), i \in \mathbb{N}^*)$, une suite de variables indépendantes de loi uniforme sur $[-1,1]^2$. On note $T = \inf\{n; U_n^2 + V_n^2 \leq 1\}$.

1. Calculer la loi de T, et donner le paramètre de sa loi.
2. Calculer et reconnaître la loi de $(U, V) = (U_T, V_T)$.
3. Montrer que T et (U, V) sont indépendants.
4. On pose $W = U^2 + V^2$. Calculer, en utilisant la transformation de Box-Muller de l'exercice III.33, la loi de (X, Y), où :

$$X = V\sqrt{2\,|\log(W)|\,/W} \quad \text{et} \quad Y = U\sqrt{2\,|\log(W)|\,/W}.$$

Cette méthode de simulation de variables aléatoires gaussiennes a l'avantage de ne pas recourir aux fonctions trigonométriques. La simulation de (U, V) se fait par rejet d'un certain nombre de simulations. On parle de méthode du rejet, voir aussi l'exercice III.35. △

Exercice III.35.
Le but de cet exercice est de présenter la méthode du rejet pour la simulation d'une variable aléatoire de densité f donnée.

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$ et soit $A \subset \mathbb{R}^d$ un ensemble mesurable tel que $\mathbb{P}(X \in A) > 0$. Soit $(X_n, n \in \mathbb{N}^*)$ des variables aléatoires indépendantes de même loi que X. On pose $T = \inf\{n \in \mathbb{N}^*; X_n \in A\}$, avec la convention $\inf \emptyset = +\infty$, et $Y = X_T$ si $T < +\infty$ et $Y = 0$ si $T = +\infty$.

1. Montrer que les variables aléatoires Y et T sont indépendantes.
2. Montrer que la loi de T est la loi géométrique de paramètre $\mathbb{P}(X \in A)$.
3. Montrer que la loi de Y est la loi conditionnelle de X sachant $\{X \in A\}$: pour tout borélien $B \subset \mathbb{R}^d$, $\mathbb{P}(Y \in B) = \mathbb{P}(X \in B | X \in A)$.

Soit h la densité d'une variable aléatoire à valeurs dans $\mathbb{R}$. On suppose qu'il existe une densité g et une constante $c > 0$ telles que $c\,h \leq g$.

Soit $(Z_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes de même loi de densité g. Soit $(U_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires de loi uniforme sur $[0, 1]$, indépendantes et indépendantes de $(Z_n, n \in \mathbb{N}^*)$. On pose $T' = \inf\{n \in \mathbb{N}^*; U_n \leq c\,h(Z_n)/g(Z_n)\}$ et $A' = \{(z, u); g(z) > 0 \text{ et } u \leq c\,h(z)/g(z)\}$.

4. Calculer $\mathbb{P}((Z_1, U_1) \in A)$.
5. Montrer que la variable aléatoire $Z_{T'}$ a pour densité f.

△

IV

Fonctions caractéristiques

De manière très générale, la fonction caractéristique d'une variable aléatoire est la transformée de Fourier de sa probabilité image. Elle prend une expression très simple pour des variables aléatoires discrètes ou à densité. Il s'agit d'un outil puissant pour étudier les lois des variables aléatoires comme on le verra au paragraphe V.6.

Les paragraphes IV.1 et IV.2 donnent la définition et quelques propriétés des fonctions caractéristiques. On calcule au paragraphe IV.3 les fonctions caractéristiques de quelques lois usuelles.

IV.1 Définitions

Avant de définir les fonctions caractéristiques, on établit une inégalité. On rappelle que si $X = (X_1, \dots, X_d)$ est une v.a. à valeurs dans $\mathbb{R}^d$ intégrable (i.e. X_i est intégrable pour $1 \leq i \leq d$), alors $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$.

Lemme IV.1. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. On note $|\cdot|$ la norme euclidienne sur $\mathbb{R}^d$. Si $\mathbb{E}[|X|] < \infty$, alors X est intégrable et on a $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.*

Démonstration. Il s'agit d'une application de l'inégalité de Jensen car la fonction $|\cdot|$ est convexe. On en donne toutefois une démonstration directe. On note $X = (X_1, \dots, X_d)$. On a $|X_i| \leq |X|$. Comme $|X|$ est intégrable, les variables X_i sont donc intégrables. Il existe un vecteur unitaire $v \in \mathbb{R}^d$ tel que $\mathbb{E}[X] = |\mathbb{E}[X]|\, v$. Il vient $\mathbb{E}[(X, v)] = (\mathbb{E}[X], v)$. Par croissance de l'espérance, on a :

$$|\mathbb{E}[X]| = \mathbb{E}[(X, v)] \leq \mathbb{E}[|(X, v)|] \leq \mathbb{E}[|X|\,|v|] = \mathbb{E}[|X|].$$

□

Soit $u \in \mathbb{R}$. La fonction complexe définie sur $\mathbb{R}$ par $x \mapsto \mathrm{e}^{iux}$ est mesurable, bornée en module par 1. Si X est une v.a. réelle, l'espérance de e^{iuX} a donc un sens grâce au lemme ci-dessus où l'on identifie l'espace complexe à $\mathbb{R}^2$.

Définition IV.2. *Soit X une v.a. réelle, la fonction complexe définie par :*

$$\boxed{\psi_X(u) = \mathbb{E}\left[\mathrm{e}^{iuX}\right],} \quad u \in \mathbb{R},$$

s'appelle la fonction ***caractéristique*** *de X.*

Remarque. Deux v.a. de même loi ont même fonction caractéristique. ◊

Exemple. Si $\mathcal{L}(X)$ est la loi uniforme sur $[a, b]$, alors $\psi_X(u) = \dfrac{\mathrm{e}^{iub} - \mathrm{e}^{iua}}{iu(b-a)}$, pour $u \in \mathbb{R}$. ◊

Si X est une v.a. à valeurs entières, la fonction caractéristique apparaît comme le prolongement de la fonction génératrice ϕ_X sur le cercle unité complexe. En effet, on a $\psi_X(u) = \phi_X\left(\mathrm{e}^{iu}\right)$.
Exemple. Si $\mathcal{L}(X)$ est la loi de Bernoulli de paramètre p, alors $\psi_X(u) = 1-p+p\,\mathrm{e}^{iu}$. ◊

Proposition IV.3. *La fonction caractéristique ψ_X de la v.a. X satisfait les conditions suivantes :*

1. *ψ_X est continue.*
2. *$|\psi_X(u)| \leq 1$.*
3. *$\psi_X(0) = 1$.*
4. *$\psi_X(-u) = \overline{\psi_X(u)}$.*

Démonstration. La propriété 1 est une conséquence directe du théorème V.3 de convergence dominée pour l'espérance. On peut cependant en donner une démonstration directe. Soit $\varepsilon > 0$ et $u \in \mathbb{R}$ fixés. Comme $\mathbb{P}(X \in \mathbb{R}) = 1$, on déduit de la propriété de convergence monotone des probabilités (cf. proposition I.2 4.) qu'il existe $n > 0$ tel que $\mathbb{P}(X \in [-n, n]) \geq 1 - (\varepsilon/3)$. La fonction $x \mapsto \mathrm{e}^{ix}$ est continue

et périodique. Donc il existe $\eta > 0$ tel que si $|x-y| \leq n\eta$, alors $\left|e^{ix} - e^{iy}\right| \leq \varepsilon/3$. On en déduit donc que pour $|u-u'| \leq \eta$, on a $\mathbf{1}_{\{X\in[-n,n]\}}\left|e^{iuX} - e^{iu'X}\right| \leq \varepsilon/3$. On remarque enfin que grâce au lemme IV.1, puis à la croissance de l'espérance, il vient :

$$\begin{aligned}\left|\psi_X(u) - \psi_X(u')\right| &\leq \mathbb{E}\left[\left|e^{iuX} - e^{iu'X}\right|\right] \\ &\leq 2\mathbb{P}(X \notin [-n,n]) + \mathbb{E}\left[\mathbf{1}_{\{X\in[-n,n]\}}\left|e^{iuX} - e^{iu'X}\right|\right] \\ &\leq \varepsilon.\end{aligned}$$

La fonction caractéristique est donc continue au point u, et ce pour tout $u \in \mathbb{R}$.

La propriété 2 découle du lemme IV.1. La propriété 3 est claire. Pour la propriété 4, on écrit $e^{iuX} = \cos(uX) + i\sin(uX)$ et on utilise la linéarité de l'espérance. □

La définition des fonctions caractéristiques se généralise au cas des v.a. vectorielles.

Définition IV.4. *Soit $X = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$. La fonction complexe définie par :*

$$\boxed{\psi_X(u) = \mathbb{E}\left[e^{i(u_1X_1+\cdots+u_dX_d)}\right],} \quad u = (u_1, \ldots, u_d) \in \mathbb{R}^d,$$

*s'appelle la fonction **caractéristique** de X.*

La proposition IV.3 est également vraie pour les v.a. vectorielles.

IV.2 Propriétés

Si X est une v.a.c. réelle de densité f, on a :

$$\psi_X(u) = \int_{\mathbb{R}} e^{iux}\, f(x)\, dx = \hat{f}(u),$$

où $\hat{f}$ est la transformée de Fourier de f. On sait dans certains cas inverser la transformation de Fourier. Ainsi si $\hat{f}$ est intégrable ($\int_{\mathbb{R}} \left|\hat{f}(u)\right|\, du < \infty$), alors on peut retrouver la fonction f à l'aide de la transformée de Fourier inverse de $\hat{f}$:

$$f(x) = \int_{\mathbb{R}} \mathrm{e}^{-iux}\, \hat{f}(u)\, \frac{du}{2\pi}, \quad \text{pour presque tout } x \in \mathbb{R}.$$

Ainsi si ψ_X est intégrable, on peut retrouver la densité de la loi à partir de la fonction caractéristique.

Le théorème suivant que l'on admet, assure qu'il y a une bijection entre les lois et les fonctions caractéristiques.

Théorème IV.5. *La fonction caractéristique caractérise la loi : deux v.a. ont même loi si et seulement si elles ont même fonction caractéristique.*

Si X et Y sont deux v.a. réelles ou vectorielles, alors les propriétés suivantes sont équivalentes :

1. X et Y ont même loi.
2. $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ pour tout borélien A.
3. $\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$ pour toute fonction g bornée mesurable.
4. Les fonctions de répartitions sont égales : $F_X = F_Y$.
5. Les fonctions caractéristiques sont égales : $\psi_X = \psi_Y$.
6. (Si X et Y sont des v.a. continues.) Les densités sont égales presque partout : $f_X = f_Y$ p.p.
7. (Si X et Y sont des v.a. discrètes.) $\mathbb{P}(X = x) = \mathbb{P}(Y = x)$ pour tout x.

L'équivalence $1 \Leftrightarrow 2$ provient de la définition de la loi, l'équivalence $2 \Leftrightarrow 3$ se déduit de la définition de l'espérance, l'équivalence $1 \Leftrightarrow 4$ correspond au théorème III.5.

Proposition IV.6.

1. *Soit $X_1, \ldots, X_n$ des v.a. réelles* ***indépendantes****, alors on a :*

$$\psi_{X_1+\cdots+X_n}(u) = \prod_{i=1}^{n} \psi_{X_i}(u), \quad \forall u \in \mathbb{R}.$$

2. *Soit $X_1, \ldots, X_n$ des v.a. réelles. Ces variables sont* ***indépendantes si et seulement si*** *pour tout $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$:*

$$\boxed{\psi_{X_1,\ldots,X_n}(u) = \prod_{i=1}^{n} \psi_{X_i}(u_i).}$$

3. *Soit $a, b \in \mathbb{R}$ et X une v.a. réelle, alors on a :*

$$\boxed{\psi_{aX+b}(u) = \mathrm{e}^{ibu}\,\psi_X(au),} \quad \forall u \in \mathbb{R}.$$

4. *Soit $X = (X_1, \ldots, X_n)$ une v.a. à valeurs dans $\mathbb{R}^n$. Soit $m \in \mathbb{N}$. On suppose que $\mathbb{E}\left[|X|^m\right] < \infty$. Alors, ψ_X possède des dérivées partielles continues d'ordre $k \leq m$, et pour tout $k_1, \ldots, k_n \in \mathbb{N}$ tel que $k = k_1 + \cdots + k_n \leq m$, on a :*

$$\frac{\partial^k \psi_X}{\partial^{k_1} u_1 \cdots \partial^{k_n} u_n}(u_1, \ldots, u_n) = i^k \mathbb{E}\left[X_1^{k_1} \cdots X_n^{k_n}\ \mathrm{e}^{i(\sum_{j=1}^n u_j X_j)}\right].$$

Les propriétés 1, 2 et 3 de la proposition IV.6 s'étendent aux v.a. vectorielles.

Démonstration. La propriété 1 est une application de la définition des v.a. indépendantes.

On montre la propriété 2. Si les v.a. sont indépendantes, on a :

$$\psi_{X_1,\ldots,X_n}(u) = \mathbb{E}[\mathrm{e}^{iu_1X_1+\cdots+iu_nX_n}] = \prod_{k=1}^n \mathbb{E}[\mathrm{e}^{iu_kX_k}] = \prod_{k=1}^n \psi_{X_k}(u_k).$$

Pour la réciproque, on suppose que pour tout $u \in \mathbb{R}^n$, on a :

$$\psi_{X_1,\ldots,X_n}(u) = \prod_{i=1}^n \psi_{X_i}(u_i).$$

Soit $(\tilde{X}_1, \ldots, \tilde{X}_n)$ une famille de variables aléatoires indépendantes telle que $\tilde{X}_i$ ait même loi que X_i pour tout $i \in \{1, \ldots, n\}$. Alors on déduit de ce qui précède que $\psi_{\tilde{X}_1,\ldots,\tilde{X}_n}(u) = \prod_{i=1}^n \psi_{\tilde{X}_i}(u_i)$. Mais comme les v.a. $\tilde{X}_1, \ldots, \tilde{X}_n$ ont même loi que $X_1, \ldots, X_n$, on a donc :

$$\psi_{\tilde{X}_1,\ldots,\tilde{X}_n}(u) = \prod_{i=1}^n \psi_{\tilde{X}_i}(u_i) = \prod_{i=1}^n \psi_{X_i}(u_i) = \psi_{X_1,\ldots,X_n}(u).$$

Donc $(\tilde{X}_1, \ldots, \tilde{X}_n)$ a même loi que $(X_1, \ldots, X_n)$. En particulier, les v.a. $X_1, \ldots, X_n$ sont indépendantes.

Enfin la propriété 3 est une conséquence de la définition des fonctions caractéristiques. On admet la propriété 4. □

Remarque. La fonction caractéristique caractérise la loi de la variable aléatoire. On peut alors se demander, à quelles conditions une fonction ψ définie sur $\mathbb{R}$ à valeurs complexes est la fonction caractéristique d'une v.a. réelle. Le théorème de Bochner assure qu'il suffit que la fonction ψ vérifie les conditions suivantes :

1. $\psi(0) = 1$.
2. ψ est continue en 0.
3. Pour toute suite finie de complexes $(\alpha_i, i \in I)$ et de réels $(u_i, i \in I)$, on a :

$$\sum_{i,j \in I} \alpha_i \overline{\alpha_j} \psi(u_i - u_j) \geq 0. \tag{IV.1}$$

$\Diamond$

Exercice IV.1.
Montrer que la condition (IV.1) est vérifiée par les fonctions caractéristiques. $\triangle$

IV.3 Fonctions caractéristiques usuelles

On donne les fonctions caractéristiques de quelques lois usuelles.

Proposition IV.7.

Bernoulli (p) :	$\psi(u) = (1-p) + p\,\mathrm{e}^{iu}$.
binomiale (n,p) :	$\psi(u) = \left[(1-p) + p\,\mathrm{e}^{iu}\right]^n$.
géométrique (p) :	$\psi(u) = p\,\mathrm{e}^{iu} \,/\, \left[1 - (1-p)\,\mathrm{e}^{iu}\right]$.
Poisson (θ) :	$\psi(u) = \mathrm{e}^{-\theta(1-\mathrm{e}^{iu})}$.
uniforme $[-1,1]$:	$\psi(u) = \dfrac{\sin(u)}{u}$.
gaussienne $\mathcal{N}(0,1)$:	$\psi(u) = \mathrm{e}^{-u^2/2}$.
gaussienne $\mathcal{N}(m,\sigma^2)$:	$\boxed{\psi(u) = \mathrm{e}^{imu - \frac{\sigma^2 u^2}{2}}}$.
exponentielle (λ) :	$\psi(u) = \dfrac{\lambda}{\lambda - iu}$.
gamma $\Gamma(\lambda,\alpha)$:	$\psi(u) = \left(\dfrac{\lambda}{\lambda - iu}\right)^{\alpha}$.
Cauchy (a) :	$\psi(u) = \mathrm{e}^{-a\,\lvert u\rvert}$.

Démonstration. Le calcul de la fonction caractéristique est immédiat pour les v.a.d. ainsi que pour la loi uniforme sur $[-1,1]$. Pour la fonction caractéristique de la loi gaussienne, grâce à la propriété 3 de la proposition IV.6 et l'exercice III.3, il suffit de calculer la fonction caractéristique de la loi $\mathcal{N}(0,1)$.

Soit X une v.a. de loi $\mathcal{N}(0,1)$. Soit $\lambda \in \mathbb{R}$. On a :

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \int_{\mathbb{R}} \mathrm{e}^{\lambda x - \frac{x^2}{2}} \ \frac{dx}{\sqrt{2\pi}} = \mathrm{e}^{\lambda^2/2} \int_{\mathbb{R}} \mathrm{e}^{-\frac{(x-\lambda)^2}{2}} \ \frac{dx}{\sqrt{2\pi}} = \mathrm{e}^{\lambda^2/2}\,.$$

On désire étendre cette égalité pour $\lambda \in \mathbb{C}$. Soit $\lambda \in \mathbb{C}$. On remarque d'abord que $\left|\mathrm{e}^{\lambda x}\right| \leq \mathrm{e}^{|\lambda|\,|x|}$. Comme $\displaystyle\int_{\mathbb{R}} \mathrm{e}^{|\lambda|\,|x| - \frac{x^2}{2}} \ \frac{dx}{\sqrt{2\pi}}$ est fini, on en déduit que $\mathrm{e}^{\lambda X}$ est intégrable. Et on a :

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \int_{\mathbb{R}} \mathrm{e}^{\lambda x - \frac{x^2}{2}} \ \frac{dx}{\sqrt{2\pi}}.$$

Pour calculer cette dernière intégrale, on introduit $g_n(x) = \sum_{k=0}^{n} \frac{(\lambda x)^k}{k!}$. On a $\lim_{n\to\infty} g_n(x) = \mathrm{e}^{\lambda x}$ et les fonctions $g_n(x)\,\mathrm{e}^{-x^2/2}$ sont toutes bornées (en module) par $h(x) = \mathrm{e}^{|\lambda|\,|x| - \frac{x^2}{2}}$ qui est intégrable. Par le théorème de convergence dominée (théorème III.25), on a :

$$\lim_{n\to\infty} \int_{\mathbb{R}} g_n(x)\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}} = \int_{\mathbb{R}} \mathrm{e}^{\lambda x - \frac{x^2}{2}} \ \frac{dx}{\sqrt{2\pi}} = \mathbb{E}\left[\mathrm{e}^{\lambda X}\right].$$

Par linéarité, on a également :

$$\int_{\mathbb{R}} g_n(x)\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}} = \sum_{k=0}^{n} \frac{\lambda^k}{k!} \int_{\mathbb{R}} x^k\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}}.$$

À l'aide d'une intégration par partie, on démontre facilement par récurrence que :

$$\mathbb{E}[X^{2m}] = \int_{\mathbb{R}} x^{2m}\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}} = \frac{(2m)!}{2^m(m!)},$$
$$\mathbb{E}[X^{2m+1}] = \int_{\mathbb{R}} x^{2m+1}\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}} = 0.$$

On en déduit que :

$$\int_{\mathbb{R}} g_n(x)\,\mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}} = \sum_{k=0}^{n} \frac{\lambda^{2k}}{2^k(k!)}.$$

Par passage à la limite, on obtient pour $\lambda \in \mathbb{C}$:

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \int_{\mathbb{R}} \mathrm{e}^{\lambda x - \frac{x^2}{2}} \ \frac{dx}{\sqrt{2\pi}} = \sum_{k=0}^{+\infty} \frac{\lambda^{2k}}{2^k(k!)} = \mathrm{e}^{\frac{\lambda^2}{2}}\,.$$

On en déduit ainsi la fonction caractéristique de la loi normale.

Le calcul de la fonction caractéristique de la loi exponentielle est immédiat. En revanche celui de la loi gamma est plus délicat. On admet ce résultat. On calcule la fonction caractéristique de la loi de Cauchy dans l'exercice suivant. □

Exercice IV.2.
Soit Y une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$ et ε une variable aléatoire indépendante de Y et telle que $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$.

1. Calculer la densité et la fonction caractéristique de $Z = \varepsilon Y$. La loi de Z est appelée loi exponentielle symétrique.
2. En déduire la fonction caractéristique de la loi de Cauchy.

△

Correction IV.2.

1. La densité de la loi de Z, f_Z, a été calculée dans l'exercice III.5 : $f_Z(z) = \dfrac{\lambda}{2}\,\mathrm{e}^{-\lambda|z|}$. On utilise la formule de décomposition et l'indépendance entre Y et ε pour obtenir :

$$\begin{aligned}\psi_Z(u) &= \mathbb{E}\left[\mathrm{e}^{iuY}\,\mathbf{1}_{\{\varepsilon=1\}}\right] + \mathbb{E}\left[\mathrm{e}^{-iuY}\,\mathbf{1}_{\{\varepsilon=-1\}}\right] \\ &= \frac{1}{2}\left[\frac{\lambda}{\lambda - iu} + \overline{\left(\frac{\lambda}{\lambda - iu}\right)}\right] = \frac{\lambda^2}{\lambda^2 + u^2}.\end{aligned}$$

2. On remarque que $\dfrac{1}{\lambda\pi}\,\psi_Z$ est la densité de la loi de Cauchy de paramètre λ. À l'aide du théorème d'inversion de la transformée de Fourier pour les fonctions intégrables, on a donc p.p. $f_Z(z) = \displaystyle\int_{\mathbb{R}} \mathrm{e}^{-iuz}\,\psi_Z(u)\,\frac{du}{2\pi}$. Comme les membres de droite et de gauche sont des fonctions continues, on a l'égalité pour tout $z \in \mathbb{R}$. On a donc $\dfrac{\lambda}{2}\,\mathrm{e}^{-\lambda|z|} = \displaystyle\int_{\mathbb{R}} \mathrm{e}^{-iuz}\,\frac{\lambda^2}{\lambda^2 + u^2}\,\frac{du}{2\pi}$. On en déduit ainsi la fonction caractéristique, ψ, de la loi de Cauchy de paramètre λ : pour tout $z \in \mathbb{R}$,

$$\psi(z) = \int_{\mathbb{R}} \mathrm{e}^{iuz}\,\frac{1}{\pi}\,\frac{\lambda}{\lambda^2 + u^2}\,du = \mathrm{e}^{-\lambda|z|}.$$

▲

IV.4 Résumé

- Soit $X = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$. Sa **fonction caractéristique** est $\psi_X(u) = \mathbb{E}\left[\mathrm{e}^{i(u_1X_1+\cdots+u_dX_d)}\right]$, où $u = (u_1, \ldots, u_d) \in \mathbb{R}^d$.
- Pour $a, b \in \mathbb{R}$ et X à valeurs réelles, on a $\psi_{aX+b}(u) = \mathrm{e}^{iub}\,\psi_X(au)$.
- Les v.a. $X_1, \ldots, X_d$ sont indépendantes si et seulement si :

$$\psi_{X_1,\ldots,X_d}(u_1, \ldots, u_d) = \prod_{i=1}^{d} \psi_{X_i}(u_i) \quad \text{pour tout} \quad (u_1, \ldots, u_d).$$

- Si les v.a. $X_1, \ldots, X_d$ sont indépendantes, alors on a :

$$\psi_{X_1+\cdots+X_d}(u) = \prod_{i=1}^{d} \psi_{X_i}(u) \quad \text{pour tout} \quad u.$$

- Les fonctions caractéristiques des lois usuelles sont :

Loi (v.a.d.)	Fonction caractéristique
Bernoulli (p)	$\psi(u) = (1-p) + p\,\mathrm{e}^{iu}$.
binomiale (n, p)	$\psi(u) = \left[(1-p) + p\,\mathrm{e}^{iu}\right]^n$.
géométrique (p)	$\psi(u) = p\,\mathrm{e}^{iu} \,/\, \left[1 - (1-p)\,\mathrm{e}^{iu}\right]$.
Poisson (θ)	$\psi(u) = \mathrm{e}^{-\theta(1-\mathrm{e}^{iu})}$.

Loi (v.a.c.)	Fonction caractéristique
uniforme $[-1,1]$	$\psi(u) = \dfrac{\sin(u)}{u}$.
gaussienne $\mathcal{N}(0,1)$	$\psi(u) = \mathrm{e}^{-u^2/2}$.
gaussienne $\mathcal{N}(m,\sigma^2)$	$\psi(u) = \mathrm{e}^{im\,u-\frac{\sigma^2 u^2}{2}}$.
exponentielle (λ)	$\psi(u) = \dfrac{\lambda}{\lambda - iu}$.
gamma $\Gamma(\lambda,\alpha)$	$\psi(u) = \left(\dfrac{\lambda}{\lambda - iu}\right)^{\alpha}$.
Cauchy (a)	$\psi(u) = \mathrm{e}^{-a\,\lvert u\rvert}$.

IV.5 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice IV.1 p. 118, *Exercice* IV.2 p. 120.

Exercice IV.3.
Soit X_1, X_2 deux v.a. indépendantes ayant pour lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$. Montrer que la loi de X_1+X_2 est la loi gaussienne $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$. △

Exercice IV.4.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. indépendantes de loi de Cauchy de paramètre a_n. Montrer que la loi de la moyenne empirique $\frac{1}{n}\sum_{i=1}^n X_i$ est une loi de Cauchy de paramètre $\frac{1}{n}\sum_{i=1}^n a_i$. En particulier, si les v.a. indépendantes suivent une loi de Cauchy de même paramètre a, alors la loi de la moyenne empirique suit la loi de Cauchy de paramètre a. △

Exercice IV.5.
Soit X_1, X_2 deux v.a. indépendantes et de lois respectives $\Gamma(\lambda, \alpha_1)$ et $\Gamma(\lambda, \alpha_2)$. Le paramètre λ est identique. Montrer que la loi de $X_1 + X_2$ est une loi gamma de paramètre $(\lambda, \alpha_1 + \alpha_2)$. En déduire que, si $(X_n, n \in \mathbb{N}^*)$ est une suite de v.a. indépendantes de loi exponentielle de paramètre $\lambda > 0$, alors la loi de la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est la loi $\Gamma(n\lambda, n)$. △

Exercice IV.6.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. indépendantes, telle que X_n est de loi $\chi^2(d_n)$. Montrer que la loi de la somme $S_n = \sum_{i=1}^n X_i$ est la loi du χ^2 de paramètre $\sum_{i=1}^n d_i$. △

Exercice IV.7.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. indépendantes et de même loi exponentielle de paramètre $\lambda > 0$. Soit T une v.a. de loi géométrique de paramètre $p \in]0, 1[$, indépendante de la suite de v.a. $(X_n, n \in \mathbb{N}^*)$. Montrer que la loi de $Z = \sum_{i=1}^T X_i$ est une loi exponentielle de paramètre $p\lambda$. △

Exercice IV.8.
Soit X une v.a. réelle dont la fonction caractéristique est $\psi_X(u)$. Montrer que $|\psi_X(u)|^2$ est la fonction caractéristique d'une v.a. réelle. On pourra écrire $|\psi_X(u)|^2$ comme le produit de deux fonctions. △

Exercice IV.9.
Soit $(T_k, k \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes de loi exponen-

tielle de paramètre $\lambda > 0$. On définit pour tout $t \geq 0$, $N_t = \inf\{k; T_1 + \ldots + T_{k+1} \geq t\}$. Le processus $(N_t, t \in \mathbb{R}^+)$ est appelé processus de Poisson de paramètre $\lambda > 0$. Ce processus permet, par exemple, de modéliser le processus d'arrivée des clients à un guichet.

1. Calculer la loi de $\Gamma_k = \sum_{i=1}^k T_i$.
2. Calculer la loi de N_t. Vérifier que $\mathbb{P}(N_t = k) = \mathrm{e}^{-\lambda t}\, \mathbb{E}[\mathbf{1}_{\{\Gamma_k \leq t\}}\, \mathrm{e}^{\lambda \Gamma_k}]$.
3. Montrer que $\mathbb{P}(N_t = k, N_{t+h} - N_t \geq l) = \mathbb{P}(N_t = k)\mathbb{P}(N_h \geq l)$.
4. En déduire que pour tout $t, h \geq 0$, $N_{t+h} - N_t$ a même loi que N_h. On dit que les accroissements sont stationnaires.
5. Vérifier que N_t et $N_{t+h} - N_t$ sont indépendants. Cette propriété d'indépendance des accroissements se généralise à un nombre quelconque d'accroissements (disjoints).

△

Exercice IV.10.
Soit $(N_t, t \in \mathbb{R}^+)$ un processus de Poisson de paramètre $\lambda > 0$ (cf l'exercice IV.9). On note $T_0 = 0$ et pour $k \geq 1$, $T_k = \inf\{t \geq 0; N_{\sum_{i=0}^{k-1} T_i + t} = k\}$.

1. Calculer la loi de $(T_1, \ldots, T_k)$ conditionnellement à $N_t = k$.
2. Montrer que conditionnellement à $N_t = k$, $(T_1, T_1 + T_2, \ldots, T_1 + \cdots + T_k)$ a même loi que le réordonnement croissant de k variables aléatoires indépendantes de loi uniforme sur $[0, t]$.

△

V

Convergences et théorèmes limites

Ce chapitre est consacré aux deux résultats fondamentaux des probabilités : la loi forte des grands nombres qui assure que la moyenne arithmétique de variables aléatoires indépendantes et de même loi converge vers une constante quand le nombre de variables aléatoires augmente, et le théorème central limite qui précise la vitesse de cette convergence.

L'énoncé de ces résultats nécessite l'introduction de plusieurs modes de convergence pour les variables aléatoires : la convergence presque sûre au paragraphe V.1, la convergence en probabilité et la convergence quadratique au paragraphe V.2, et la convergence en loi au paragraphe V.3.

La loi forte des grands nombres est énoncée et en partie démontrée au paragraphe V.4. On donne des applications de ce résultat au paragraphe V.5. Le théorème central limite est énoncé et démontré au paragraphe V.6. La notion d'intervalle de confiance, qui par exemple permet de donner la validité d'un sondage, est introduite au paragraphe V.7. D'autres résultats qui complètent le théorème central limite sont présentés dans le paragraphe V.8.

V.1 Convergence presque sûre et théorèmes limites

On présente dans ce paragraphe la convergence presque sûre (p.s.) et les conditions qui permettent de permuter convergence p.s. et espérance. Les définitions de la limite supérieure et de la limite inférieure sont rappelées au paragraphe I.10.

Définition V.1. *On dit qu'une suite $(X_n, n \in \mathbb{N})$ de variables aléatoires réelles définies sur le même espace probabilisé,* ***converge presque sûrement*** *si :*

$$\mathbb{P}\left(\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n\right) = 1.$$

On note $\lim_{n\to\infty} X_n$ la v.a. limite. Une suite de variables aléatoires vectorielles $((X_n(1), \ldots, X_n(d)), n \in \mathbb{N})$ converge p.s. si les suites des coordonnées $(X_n(i), n \in \mathbb{N})$, pour $i \in \{1, \ldots, d\}$, convergent presque sûrement. Enfin si la suite $(X_n, n \in \mathbb{N}^*)$ converge p.s. vers X, on le note :

$$X_n \xrightarrow[n\to\infty]{\textbf{p.s.}} X.$$

Remarque. La limite, quand elle existe, est mesurable (cf. la proposition I.19), c'est donc une variable aléatoire. ◊

Exemple. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. de loi de Bernoulli de paramètre p_n. Alors la suite des sommes partielles $(\sum_{k=0}^{n} 2^{-k} X_k, n \in \mathbb{N}^*)$ converge p.s. ◊

Remarque. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. indépendantes de loi uniforme sur $[-1, 1]$. La suite des sommes partielles $(\sum_{k=1}^{n} X_k/k^2, n \in \mathbb{N}^*)$ converge (absolument) p.s. Le graphe V.1 représente des réalisations des ces sommes partielles. ◊

Lemme V.2. *Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. qui converge p.s. vers X. Soit h une fonction mesurable. Soit C l'ensemble des points où la fonction h est continue. Si $\mathbb{P}(X \in C) = 1$, alors la suite de v.a. $(h(X_n), n \in \mathbb{N})$ converge p.s. vers $h(X)$.*

Démonstration. Soit l'évènement $A = \{X_n \text{ converge vers } X\}$. Par hypothèse on a $\mathbb{P}(A) = 1$. Pour $\omega \in A \cap \{X \in C\}$, on a $\lim_{n\to\infty} h(X_n(\omega)) = h(X(\omega))$. Cela signifie que :

$$A \cap \{X \in C\} \subset \{h(X_n) \text{ converge vers } h(X)\}.$$

Comme $\mathbb{P}(A \cap \{X \in C\}) = 1$, on en déduit par croissance, que la probabilité du membre de droite est 1. La suite de v.a. $(h(X_n), n \in \mathbb{N})$ converge p.s. vers $h(X)$. □

Exemple. Si $(X_n, n \in \mathbb{N}^*)$ est une suite de v.a. qui converge p.s. vers X et si $\mathbb{P}(X = 0) = 0$, alors la suite $(X_n^{-1}, n \in \mathbb{N}^*)$ converge p.s. vers X^{-1}. ◊

Les trois théorèmes suivants que l'on admet sont à rapprocher des théorèmes énoncés au paragraphe III.10 sur l'intégrale de Lebesgue.

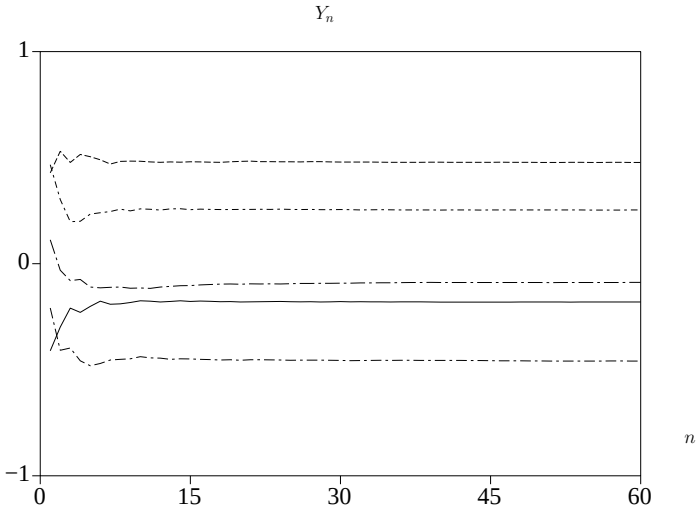


Figure V.1. 5 réalisations de la suite $n \mapsto Y_n = \sum_{k=1}^{n} X_k/k^2$.

Théorème V.3 (Convergence dominée)**.** *Soit Y une v.a. réelle positive telle que $\mathbb{E}[Y] < \infty$. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. réelles ou vectorielles telles que pour tout $n \in \mathbb{N}$ p.s. $|X_n| \leq Y$ (les v.a. X_n sont dominées p.s. par la v.a. Y et donc intégrables). Si la suite de v.a. $(X_n, n \in \mathbb{N})$ converge p.s., alors la v.a. $\lim_{n\to\infty} X_n$ est aussi intégrable, et on a :*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

Exemple V.4. On termine la démonstration de la formule de décomposition (II.10). Soit X une v.a. et Y une v.a.d. de support Δ. Soit φ une fonction mesurable bornée. On désire montrer que :

$$\mathbb{E}[\varphi(X,Y)] = \sum_{y\in\Delta} \mathbb{E}[\varphi(X,y), Y = y],$$

où par convention $\mathbb{E}\left[\varphi(X,y), Y = y\right] = \mathbb{E}\left[\varphi(X,y)\mathbf{1}_{\{Y=y\}}\right]$. On considère une suite croissante $(\Delta_n, n \in \mathbb{N})$ de sous-ensembles finis telle que $\bigcup_{n\geq 1} \Delta_n = \Delta$. On pose $Z_n = \sum_{y\in\Delta_n} \varphi(X,y)\mathbf{1}_{\{Y=y\}}$. La v.a. Z_n est dominée par $\|\varphi\|_\infty \sum_{y\in\Delta} \mathbf{1}_{\{Y=y\}} \leq \|\varphi\|_\infty$. De plus la suite $(Z_n, n \in \mathbb{N})$ converge p.s. vers $\sum_{y\in\Delta} \varphi(X,y)\mathbf{1}_{\{Y=y\}} = \varphi(X,Y)$. Par le théorème de convergence dominée, on obtient :

$$\sum_{y\in\Delta}\mathbb{E}[\varphi(X,y),Y=y]=\lim_{n\to\infty}\mathbb{E}[Z_n]=\mathbb{E}\left[\lim_{n\to\infty}Z_n\right]=\mathbb{E}[\varphi(X,Y)].$$

$\Diamond$

Théorème V.5 (Convergence monotone)**.** *Soit $(X_n, n\in\mathbb{N})$ une suite croissante de v.a. réelles* ***positives*** *: pour tout $m\geq n\geq 0$, p.s. $X_m\geq X_n\geq 0$. On a :*

$$\lim_{n\to\infty}\mathbb{E}[X_n]=\mathbb{E}\left[\lim_{n\to\infty}X_n\right],$$

où les limites peuvent éventuellement prendre la valeur $+\infty$.

Théorème V.6 (Lemme de Fatou)**.** *Soit $(X_n, n\in\mathbb{N})$ une suite de v.a.* ***positives****. On a :*

$$\liminf_{n\to\infty}\mathbb{E}[X_n]\geq\mathbb{E}\left[\liminf_{n\to\infty}X_n\right].$$

V.2 Convergence en probabilité et dans l'espace L^2

On présente dans ce paragraphe la convergence en probabilité et dans L^2.

Définition V.7. *On dit qu'une suite de v.a. $(X_n, n\in\mathbb{N})$ définies sur le même espace probabilisé* ***converge en probabilité*** *vers une v.a. X si : pour tout $\varepsilon>0$,*

$$\boxed{\lim_{n\to\infty}\mathbb{P}(|X_n-X|>\varepsilon)=0.}$$

On a déjà vu la convergence en probabilité au paragraphe II.11 pour la loi faible des grands nombres.

Exemple. Soit $(X_n, n\in\mathbb{N})$ une suite de v.a. de Bernoulli de paramètre p_n tel que $\lim_{n\to\infty}p_n=0$. Alors la suite converge en probabilité vers 0. En effet, pour $\varepsilon\in]0,1[$, on a $\mathbb{P}(|X_n|>\varepsilon)=p_n$ et $\lim_{n\to\infty}p_n=0$. $\Diamond$

Proposition V.8. *La convergence p.s. entraîne la convergence en probabilité.*

Démonstration. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a qui converge p.s. vers X. La suite de v.a. discrètes positives $(\mathbf{1}_{\{|X_n - X| > \varepsilon\}}, n \in \mathbb{N})$ converge p.s. vers 0. De plus elle est uniformément bornée par 1. Par le théorème de convergence dominée, on en déduit que :

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = \lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_{\{|X_n - X| > \varepsilon\}}\right] = \mathbb{E}\left[\lim_{n\to\infty} \mathbf{1}_{\{|X_n - X| > \varepsilon\}}\right] = 0.$$

□

Remarque. La **réciproque est fausse en général** comme le montre l'exemple suivant. Soit l'espace probabilisé $([0,1], \mathcal{B}([0,1]), \lambda)$, où λ est la mesure de Lebesgue sur $[0,1]$. Les v.a. réelles sont alors les fonctions réelles mesurables définies sur $[0,1]$. On définit les v.a. X_{2^n+k} de la manière suivante : pour $n \in \mathbb{N}$ et $k \in \{0, \cdots, 2^n - 1\}$,

$$X_{2^n+k}(\omega) = \mathbf{1}_{[k2^{-n}, (k+1)2^{-n}]}(\omega).$$

Pour tout $\omega \in [0,1]$, la suite $X_p(\omega)$ prend une infinité de fois les valeurs 0 et 1. En particulier on a $\limsup_{p\to\infty} X_p(\omega) = 1$ et $\liminf_{p\to\infty} X_p(\omega) = 0$. P.s. la suite ne converge pas. En revanche, pour $\varepsilon \in]0,1[$, on a $\mathbb{P}(|X_{2^n+k}| > \varepsilon) = \mathbb{P}(X_{2^n+k} = 1) = 2^{-n}$. Cela implique que $\lim_{p\to\infty} \mathbb{P}(|X_p| > \varepsilon) = 0$. Donc la suite converge en probabilité vers 0. ◊

On pourra dans une première lecture omettre la fin de ce paragraphe sur la réciproque partielle de la proposition V.8 et la notion de convergence dans L^2.

Proposition V.9. *De toute suite de v.a. qui converge en probabilité, on peut extraire une sous-suite qui converge presque sûrement.*

Démonstration. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. qui converge en probabilité vers X. On définit la sous-suite $(X_{\sigma(n)}, n \in \mathbb{N})$ de la manière suivante : $\sigma(0) = 0$ et pour $n \in \mathbb{N}$,

$$\sigma(n+1) = \inf\left\{p > \sigma(n) \text{ tel que } \mathbb{P}\left(|X_p - X| > \frac{1}{n+1}\right) \leq \frac{1}{(n+1)^2}\right\}.$$

La suite $(X_n, n \in \mathbb{N})$ converge en probabilité, cela assure que la sous-suite $(\sigma(n), n \in \mathbb{N})$ est bien définie. On en déduit, par convergence monotone, que :

$$\mathbb{E}\left[\sum_{n\geq 1} \mathbf{1}_{\left\{|X_{\sigma(n)} - X| > \frac{1}{n}\right\}}\right] = \sum_{n\geq 1} \mathbb{P}\left(|X_{\sigma(n)} - X| > \frac{1}{n}\right) \leq \sum_{n\geq 1} \frac{1}{n^2} < \infty.$$

Cela implique que la v.a. $\sum_{n\geq 1} \mathbf{1}_{\{|X_{\sigma(n)}-X|>\frac{1}{n}\}}$ est intégrable. En particulier, elle est finie p.s. Les termes d'une série convergente tendent vers 0. Donc p.s., on a $\lim_{n\to\infty} \mathbf{1}_{\{|X_{\sigma(n)}-X|>\frac{1}{n}\}} = 0$. Comme la fonction indicatrice ne prend que deux valeurs 0 ou 1, cela entraîne que p.s. $\mathbf{1}_{\{|X_{\sigma(n)}-X|>\frac{1}{n}\}}(\omega)$ est nul à partir d'un certain rang n_0 (qui dépend de ω). Donc, p.s. à partir d'un certain rang, on a $\left|X_{\sigma(n)}(\omega) - X(\omega)\right| \leq \frac{1}{n}$. En particulier, cela implique que p.s. $\lim_{n\to\infty} X_{\sigma(n)} = X$. Donc la sous-suite $(X_{\sigma(n)}, n \in \mathbb{N})$ converge p.s. vers X. □

On définit l'espace $\mathcal{L}^2(\Omega)$, noté $\mathcal{L}^2$, comme l'espace des v.a. réelles de carré intégrable définies sur $(\Omega, \mathcal{F})$. Si on a $X \in \mathcal{L}^2$, alors $\alpha X \in \mathcal{L}^2$, où $\alpha \in \mathbb{R}$, par linéarité de l'espérance. Enfin si $X, Y \in \mathcal{L}^2$, alors comme $(X+Y)^2 \leq 2X^2 + 2Y^2$, on en déduit que $X+Y \in \mathcal{L}^2$. Ainsi $\mathcal{L}^2$ est un espace vectoriel. On note L^2 l'espace $\mathcal{L}^2$ quotienté par la relation d'équivalence définie par l'égalité p.s. (i.e. X et Y sont en relation si $X = Y$ p.s.). Ainsi $X \in L^2$ désigne un représentant de la classe $\{Y \in \mathcal{L}^2; X = Y \quad \text{p.s.}\}$.

Par l'inégalité de Cauchy-Schwarz, on remarque que si $X, Y \in L^2$, alors XY est intégrable. On vérifie facilement que $\mathbb{E}[XY]$ définit une forme bilinéaire symétrique positive sur L^2. Enfin, on déduit de l'inégalité de Tchebychev, que si $\mathbb{P}(|X| > \varepsilon) > 0$, alors $\mathbb{E}[X^2] > 0$. Par contraposée, si $\mathbb{E}[XX] = 0$, alors $X = 0$ p.s. Ainsi l'application $(X, Y) \mapsto \mathbb{E}[XY]$ définit un produit scalaire sur L^2. La norme associée est $\|X\| = \sqrt{\mathbb{E}[X^2]}$.

Définition V.10. *On dit qu'une suite de v.a. $(X_n, n \in \mathbb{N})$ de carré intégrable* ***converge dans*** L^2 *vers une v.a. de carré intégrable X si :*

$$\boxed{\lim_{n\to\infty} \mathbb{E}\left[(X_n - X)^2\right] = 0.}$$

Proposition V.11. *La convergence L^2 entraîne la convergence en probabilité.*

Démonstration. Cela découle de l'inégalité de Tchebychev. □

Remarque. La réciproque est fausse en général, car pour la convergence en probabilité, on n'impose pas que X_n soit de carré intégrable. ◊

On dit que la suite de v.a. $(X_n, n \in \mathbb{N})$ est une **suite de Cauchy** dans L^2 si pour tout $\varepsilon > 0$, il existe N tel que pour tout $n \geq N, m \geq N$, on a $\mathbb{E}\left[(X_n - X_m)^2\right] \leq \varepsilon$.

La proposition suivante assure que l'espace L^2 est **complet**. Ainsi l'espace L^2 muni du produit scalaire $(X, Y) = \mathbb{E}[XY]$ est un espace de **Hilbert**.

Proposition V.12. *Toute suite de Cauchy dans L^2 est une suite convergente dans L^2.*

Démonstration. On suppose que $(X_n, n \in \mathbb{N})$ est une suite de Cauchy dans L^2. On extrait une sous-suite de la manière suivante : $\sigma(0) = 0$ et pour tout $n \in \mathbb{N}$, on pose :

$$\sigma(n+1) = \inf\left\{k > \sigma(n); \mathbb{E}\left[(X_l - X_m)^2\right] \leq 2^{-(n+1)} \text{ pour tous } l \geq k, m \geq k\right\}.$$

En particulier, on a pour tout $n \in \mathbb{N}$, $\mathbb{E}[(X_{\sigma(n+1)} - X_{\sigma(n)})^2] \leq 2^{-n}$. Par le théorème de convergence monotone et l'inégalité de Jensen appliquée à la fonction $\varphi(x) = x^2$, on a :

$$\begin{aligned}\mathbb{E}\left[\sum_{n\in\mathbb{N}} \left|X_{\sigma(n+1)} - X_{\sigma(n)}\right|\right] &= \sum_{n\in\mathbb{N}} \mathbb{E}\left[\left|X_{\sigma(n+1)} - X_{\sigma(n)}\right|\right] \\ &\leq \sum_{n\in\mathbb{N}} \mathbb{E}\left[(X_{\sigma(n+1)} - X_{\sigma(n)})^2\right]^{1/2} < \infty.\end{aligned}$$

Cela implique que la v.a. $\sum_{n\in\mathbb{N}} \left|X_{\sigma(n+1)} - X_{\sigma(n)}\right|$ est intégrable. Elle est donc finie p.s. La série de terme général $X_{\sigma(n+1)} - X_{\sigma(n)}$ est p.s. absolument convergente. En regardant la convergence des sommes partielles, on en déduit que la suite $(X_{\sigma(n)}, n \in \mathbb{N})$ converge p.s. On note X la v.a. limite. On déduit alors du lemme de Fatou que :

$$\liminf_{n\to\infty} \mathbb{E}\left[(X_m - X_{\sigma(n)})^2\right] \geq \mathbb{E}\left[(X_m - X)^2\right].$$

En particulier, on en déduit que pour tout $\varepsilon > 0$, il existe N tel que pour tout entier $m \geq N$, $\mathbb{E}\left[(X_m - X)^2\right] \leq \varepsilon$. La suite $(X_n, n \in \mathbb{N})$ converge donc dans L^2 vers X. □

Remarque. On a également montré dans la démonstration ci-dessus, que de toute suite convergente dans L^2, on peut extraire une sous-suite qui converge presque sûrement. ◊

V.3 Convergence en loi

Parmi les notions de convergence, la convergence en loi est la convergence la plus faible. La définition suivante est valable pour les v.a. réelles et vectorielles.

Définition V.13. *On dit qu'une suite de v.a.* $(X_n, n \in \mathbb{N})$ ***converge en loi*** *vers la loi d'une v.a.* X *(par abus de langage, on dit aussi que la suite converge en loi vers* X*) si pour toute fonction* g *à valeurs réelles,* ***bornée et continue****, on a :*

$$\boxed{\lim_{n\to\infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right].}$$

On le note :

$$X_n \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} X.$$

Remarque. Le choix de fonctions continues est essentiel (voir la proposition V.19 et l'exemple V.20). La convergence en loi n'entraîne pas en général la convergence en probabilité, ni la convergence p.s., ni dans l'espace L^2. En particulier, dans la définition de la convergence en loi, on n'impose pas que les v.a. soient définies sur le même espace probabilisé. ◊

Exemple. La suite $(X_n, n \in \mathbb{N}^*)$, où X_n est de loi uniforme sur $\left\{0, \frac{1}{n}, \cdots, \frac{n-1}{n}\right\}$, converge en loi vers U de loi uniforme sur $[0,1]$. En effet, si g est continue bornée, on déduit de la convergence des sommes de Riemann que :

$$\mathbb{E}[g(X_n)] = \frac{1}{n}\sum_{k=0}^{n-1} g\left(\frac{k}{n}\right) \underset{n\to\infty}{\longrightarrow} \int_{[0,1]} g(x)\, dx = \mathbb{E}[g(U)].$$

◊

Exercice V.1.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires de loi exponentielle de paramètre λ_n. Étudier la convergence en loi dans les trois cas suivants :

1. $\lim_{n\to\infty} \lambda_n = \lambda \in]0, \infty[$,
2. $\lim_{n\to\infty} \lambda_n = +\infty$,
3. $\lim_{n\to\infty} \lambda_n = 0$.

△

Correction V.1. Soit g continue bornée.

1. On a $\mathbb{E}[g(X_n)] = \int_0^\infty \lambda_n \, \mathrm{e}^{-\lambda_n x} \, g(x) \, dx$. Il existe $n_0 \in \mathbb{N}^*$, et $0 < \lambda_- < \lambda_+ < \infty$ tels que pour tout $n \geq n_0$, on a $\lambda_n \in [\lambda_-, \lambda_+]$. On a alors $\left|\lambda_n \, \mathrm{e}^{-\lambda_n x} \, g(x)\right| \leq \|g\|_\infty \, \lambda_+ \, \mathrm{e}^{-\lambda_- x} = h(x)$. La fonction h est intégrable sur $[0, \infty[$. On a $\lim_{n\to\infty} \lambda_n \, \mathrm{e}^{-\lambda_n x} \, g(x) = \lambda \, \mathrm{e}^{-\lambda x} \, g(x)$. On déduit du théorème de convergence dominée que :

$$\mathbb{E}[g(X_n)] \underset{n\to\infty}{\longrightarrow} \int_0^\infty \lambda \, \mathrm{e}^{-\lambda x} \, g(x) \, dx.$$

   Donc la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers la loi exponentielle de paramètre λ.

2. On a $\mathbb{E}[g(X_n)] = \int_0^\infty \mathrm{e}^{-x} \, g(x/\lambda_n) \, dx$. On a la majoration $|\mathrm{e}^{-x} \, g(x/\lambda_n)| \leq \|g\|_\infty \, \mathrm{e}^{-x} = h(x)$, et la fonction h est intégrable sur $[0, \infty[$. Comme la fonction g est continue, on a $\lim_{n\to\infty} g(x/\lambda_n) = g(0)$. Par convergence dominée, il vient :

$$\mathbb{E}[g(X_n)] \underset{n\to\infty}{\longrightarrow} g(0) = \mathbb{E}[g(X)],$$

   où p.s. $X = 0$. Donc la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers 0.

3. Si la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers une variable aléatoire X, alors les fonctions caractéristiques $\psi_{X_n}(u)$ convergent vers $\psi_X(u)$ pour tout $u \in \mathbb{R}$. On a :

$$\lim_{n\to\infty} \psi_{X_n}(u) = \lim_{n\to\infty} \frac{\lambda_n}{\lambda_n - iu} = \mathbf{1}_{\{u=0\}}.$$

   La fonction $u \mapsto \mathbf{1}_{\{u=0\}}$ n'est pas continue en 0. Or les fonctions caractéristiques sont continues. Par contraposée, ce n'est donc pas la fonction caractéristique d'une variable aléatoire et la suite $(X_n, n \in \mathbb{N}^*)$ ne converge pas en loi.

▲

En fait les fonctions caractéristiques jouent un rôle important pour la convergence en loi. On admet le théorème suivant.

Théorème V.14 (Lévy)**.** *Si $(X_n, n \in \mathbb{N})$ est une suite de v.a. réelles ou vectorielles telle que $\psi_{X_n} \underset{n\to\infty}{\longrightarrow} \psi$ et si ψ est continue en 0, alors ψ est la fonction caractéristique d'une v.a. X. De plus la suite $(X_n, n \in \mathbb{N})$ converge en loi vers la loi de X.*

On en déduit le théorème suivant utile en pratique.

Théorème V.15. *Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. réelles (resp. vectorielles à valeurs dans $\mathbb{R}^d$). La suite converge en loi vers la loi d'une v.a. X si et seulement si :*

$$\boxed{\psi_{X_n}(u) \underset{n\to\infty}{\longrightarrow} \psi_X(u),} \quad \forall u \in \mathbb{R} \quad \textit{(resp. } \forall u \in \mathbb{R}^d\textit{)} \ .$$

Exercice V.2.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. indépendantes où la loi de X_n est la loi de Cauchy de paramètre $a_n > 0$. Montrer que la suite $(\sum_{i=1}^n X_i, n \in \mathbb{N}^*)$ converge en loi si et seulement si $\sum_{i=1}^n a_i$ converge vers une limite finie. Alors la loi limite est une loi de Cauchy de paramètre $\sum_{i\geq 1} a_i$. △

Proposition V.16. *La convergence en probabilité implique la convergence en loi.*

La réciproque est fausse en général, comme le montre le contre-exemple V.17 (voir l'exercice V.3 pour une réciproque partielle).

Démonstration. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. à valeurs dans $\mathbb{R}^d$, qui converge en probabilité vers une v.a. X. On a :

$$|\psi_{X_n}(u) - \psi_X(u)| = \left|\mathbb{E}\left[\mathrm{e}^{i(u,X_n)} - \mathrm{e}^{i(u,X)}\right]\right|.$$

Grâce au lemme IV.1, il vient, pour $\varepsilon > 0$:

$$\begin{aligned}
|\psi_{X_n}(u) - \psi_X(u)| &\leq \mathbb{E}\left[\left|\mathrm{e}^{i(u,X_n)} - \mathrm{e}^{i(u,X)}\right|\right] \\
&= \mathbb{E}\left[\left|1 - \mathrm{e}^{i(u,(X-X_n))}\right|\right] \\
&\leq 2\mathbb{E}\left[\mathbf{1}_{|X-X_n|\geq\varepsilon}\right] + \sup_{|x-y|<\varepsilon}\left|1 - \mathrm{e}^{i(u,(x-y))}\right|.
\end{aligned}$$

On en déduit que $\limsup_{n\to\infty} |\psi_{X_n}(u) - \psi_X(u)| \leq \sup_{|z|<\varepsilon}\left|1 - \mathrm{e}^{i(u,z)}\right|$. Comme ceci est vrai pour tout $\varepsilon > 0$, on a donc $\lim_{n\to\infty} |\psi_{X_n}(u) - \psi_X(u)| = 0$. La suite converge donc en loi. □

Exemple V.17. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. indépendantes de loi exponentielle de paramètre $\lambda = 1$. Comme les v.a. ont même loi, on a $\mathbb{E}[g(X_n)] = \mathbb{E}[g(X_1)]$ pour toute fonction g continue bornée. La suite converge donc en loi. On montre que la suite ne converge pas en probabilité. Soit $\varepsilon > 0$. On a pour $m \neq n$:

$$\mathbb{P}(|X_n - X_m| > \varepsilon) = \int_{\mathbb{R}_+^2} \mathrm{e}^{-x-y} \mathbf{1}_{\{|x-y|>\varepsilon\}} \, dxdy > 0.$$

Cette quantité est indépendante de m et n. On raisonne maintenant par l'absurde. On suppose que la suite $(X_n, n \in \mathbb{N})$ converge en probabilité vers une v.a. X. Comme :

$$\{|X_n - X_m| > \varepsilon\} \subset \{|X_n - X| > \varepsilon/2\} \cup \{|X - X_m| > \varepsilon/2\},$$

cela implique que :

$$\mathbb{P}(|X_n - X_m| > \varepsilon) \leq \mathbb{P}(|X - X_m| > \varepsilon/2) + \mathbb{P}(|X_n - X| > \varepsilon/2).$$

En particulier $\mathbb{P}(|X_n - X_m| > \varepsilon)$ converge vers 0 quand $n \to \infty$ et $m \to \infty$. Ce qui est absurde, car cette quantité strictement positive est indépendante de m et n (pour $m \neq n$). ◊

Remarque. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. qui converge en loi vers X et $(Y_n, n \in \mathbb{N})$ une suite de v.a. qui converge en loi vers Y. Alors on n'a pas forcément la convergence en loi de la suite $(X_n + Y_n, n \in \mathbb{N})$ vers $X + Y$, ni de manière plus générale celle de $((X_n, Y_n), n \in \mathbb{N})$ vers la loi du couple (X, Y). La convergence des lois marginales en loi n'implique pas la convergence en loi du vecteur, voir l'exemple V.18. ◊

Exemple V.18. Soit X une v.a. de loi $\mathcal{N}(0,1)$. Pour $n \in \mathbb{N}$, on pose $X_n = X$ et $Y_n = (-1)^n X$. Comme X et $-X$ ont même loi, on en déduit que les lois de X_n et de Y_n sont indépendantes de n, il s'agit de la loi $\mathcal{N}(0,1)$. En revanche $X_{2n} + Y_{2n} = 2X$ et $X_{2n+1} + Y_{2n+1} = 0$. La suite $(X_n + Y_n, n \in \mathbb{N})$ ne converge donc pas en loi. En utilisant les fonctions caractéristiques, on vérifie facilement que l'on n'a pas non plus la convergence de la suite $((X_n, Y_n), n \in \mathbb{N})$. La convergence en loi n'est pas une convergence d'espace vectoriel. ◊

On admet le résultat suivant qui généralise la définition V.13.

Proposition V.19. *Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. à valeurs dans $\mathbb{R}^d$ qui converge en loi vers la loi d'une v.a X. Soit h une fonction définie sur $\mathbb{R}^d$ à valeurs réelles, mesurable bornée. Soit C l'ensemble des points où la fonction est continue. Si $\mathbb{P}(X \in C) = 1$, alors on a :*

$$\lim_{n\to\infty} \mathbb{E}\left[h(X_n)\right] = \mathbb{E}\left[h(X)\right].$$

Exemple V.20. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. telles que $\mathbb{P}(X_n = 1/n) = 1$. Cette suite converge en loi vers la loi de la v.a. constante $X = 0$. Mais on a :

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_{]-\infty,0]}(X_n)\right] = 0 \neq 1 = \mathbb{E}\left[\mathbf{1}_{]-\infty,0]}(X)\right].$$

$\Diamond$

Exercice V.3.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. définies sur le même espace probabilisé, à valeurs dans $\mathbb{R}^d$ et qui converge en loi vers une constante c. Montrer que la suite converge également en probabilité vers c. $\triangle$

Correction V.3. Soit $\varepsilon > 0$. On déduit de la proposition V.19 avec la fonction $h(x) = \mathbf{1}_{\{|x-c|>\varepsilon\}}$ qui est continue en c que $\lim_{n\to+\infty} \mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(|c - c| > \varepsilon) = 0$. La suite converge donc en probabilité vers c. $\blacktriangle$

On déduit de la proposition V.19 un résultat similaire au lemme V.2 pour la convergence en loi.

Corollaire V.21. *Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. qui converge en loi vers X. Soit h une fonction mesurable. Soit C l'ensemble des points où la fonction est continue. Si $\mathbb{P}(X \in C) = 1$, alors la suite de v.a. $(h(X_n), n \in \mathbb{N})$ converge en loi vers $h(X)$.*

Démonstration. Soit φ une fonction continue bornée. Les points de continuité de $\varphi \circ h$ sont ceux de h, c'est-à-dire l'ensemble C. Comme $\mathbb{P}(X \in C) = 1$, on déduit de la proposition V.19 que $\lim_{n\to\infty} \mathbb{E}\left[\varphi(h(X_n))\right] = \mathbb{E}\left[\varphi(h(X))\right]$. Cela signifie bien que la suite de v.a. $(h(X_n), n \in \mathbb{N})$ converge en loi vers $h(X)$. $\square$

Soit X une v.a. réelle. Sa fonction de répartition $F(x) = \mathbb{P}(X \leq x) = \mathbb{E}\left[\mathbf{1}_{\{X\leq x\}}\right]$ est croissante, continue à droite et à valeurs dans $[0, 1]$. Elle possède donc au plus un nombre dénombrable de points de discontinuité. Si x est un point de continuité, alors :

$$\mathbb{P}(X = x) = \lim_{\varepsilon\to 0} \mathbb{P}(X \in]x - \varepsilon, x + \varepsilon]) = \lim_{\varepsilon\to 0} F(x + \varepsilon) - F(x - \varepsilon) = 0.$$

On déduit de la proposition V.19 que si la suite de v.a. réelles $(X_n, n \in \mathbb{N})$ converge en loi vers X, alors $\lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_{\{X_n\leq x\}}\right] = \mathbb{E}\left[\mathbf{1}_{\{X\leq x\}}\right]$. Ainsi la suite $(F_n(x), n \in \mathbb{N})$, où F_n est la fonction de répartition de X_n, converge vers $F(x)$. L'extension au cas vectoriel est immédiat.

Corollaire V.22. *Si la suite $(X_n, n \in \mathbb{N})$ de v.a. réelles ou vectorielles converge en loi vers la loi d'une variable aléatoire X, alors la suite des fonctions de répartition $(F_n, n \in \mathbb{N})$ converge ponctuellement vers la fonction de répartition, F, de la v.a. X, sauf peut-être aux points de discontinuité de F.*

On adopte la convention suivante : soit h une fonction réelle définie sur $\mathbb{R}^d$, on note :

$$\lim_{x\to-\infty} h(x) = \lim_{\max(x_1,\dots,x_n)\to-\infty} h(x_1,\dots,x_n),$$
$$\lim_{x\to+\infty} h(x) = \lim_{\min(x_1,\dots,x_n)\to+\infty} h(x_1,\dots,x_n),$$

quand les limites existent.

On admettra la réciproque suivante.

Proposition V.23. *Soit $(X_n, n \in \mathbb{N})$, une suite de v.a. réelles ou vectorielles, telle que la suite des fonctions de répartition $(F_n, n \in \mathbb{N})$ converge ponctuellement vers une fonction continue à droite F, sauf peut-être aux points de discontinuité de F. On suppose de plus que $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$. Alors la fonction F est la fonction de répartition d'une v.a. X, et la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers X.*

Exemple. Si on reprend l'exercice V.1 en utilisant les fonctions de répartition, on obtient que la limite de $F_n(x) = \mathbb{P}(X_n \le x) = (1 - \mathrm{e}^{-\lambda_n x})\mathbf{1}_{\{x\ge 0\}}$ est

- dans le cas $\lambda_n \to \lambda \in]0, +\infty[$, $(1 - \mathrm{e}^{-\lambda x})\mathbf{1}_{\{x\ge 0\}}$ qui est la fonction de répartition de la loi exponentielle de paramètre λ,
- dans le cas $\lambda_n \to +\infty$, $\mathbf{1}_{\{x\ge 0\}}$ qui est la fonction de répartition de la variable constante égale à 0,
- dans le cas $\lambda_n \to 0$, $F(x) = 0$. Cette fonction n'est pas une fonction de répartition car $\lim_{x\to+\infty} F(x) \neq 1$. Dans ce dernier cas, on retrouve bien que la suite $(X_n, n \in \mathbb{N}^*)$ ne converge pas en loi.

$\Diamond$

V.4 Loi forte des grands nombres

On améliore le résultat de la loi faible des grands nombres démontrée au paragraphe II.11 : on a en fait, sous certaines hypothèses, la convergence presque sûre,

et non pas seulement en probabilité, de la moyenne empirique vers la moyenne. C'est la loi forte des grands nombres (LFGN). Elle est illustrée sur la figure V.2.

On dit que des variables aléatoires sont **identiquement distribuées** (ou équidistribuées) si elles ont même loi ; et on utilise l'acronyme **i.i.d** pour indépendantes et identiquement distribuées.

Théorème V.24 (Loi forte des grands nombres)**.** *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de v.a. réelles ou vectorielles* ***i.i.d. et intégrables*** *(c'est-à-dire indépendantes, de même loi et telles que* $\mathbb{E}[|X_n|] < \infty$*). Alors la moyenne empirique* $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ *converge* ***presque sûrement*** *vers* $\mathbb{E}[X_1]$*. Ainsi on a :*

$$\boxed{\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\textbf{\textit{p.s.}}} \mathbb{E}[X_1].}$$

On a de plus que $\lim_{n\to\infty} \mathbb{E}\left[\left|\bar{X}_n - \mathbb{E}[X_1]\right|\right] = 0.$

En particulier ce théorème justifie l'approximation faite au paragraphe I.2 de la probabilité d'un évènement par sa fréquence empirique. Enfin ce théorème est robuste : on obtient des résultats similaires sur la convergence de la moyenne empirique sous des hypothèses plus faibles que celles du théorème.

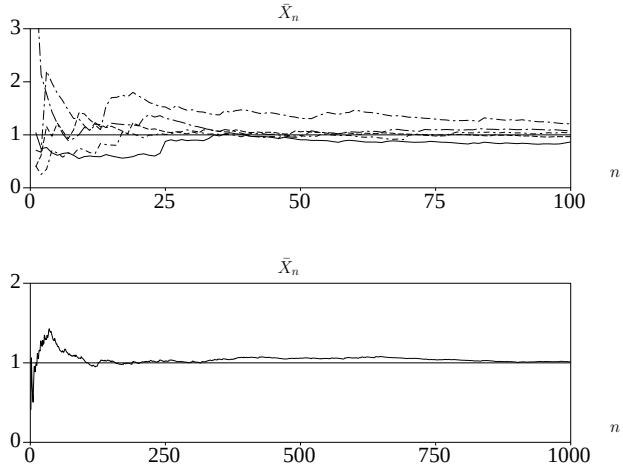


Figure V.2. Plusieurs réalisations de la suite $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, où les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes de loi exponentielle de paramètre $\lambda = 1$ (on a alors $\mathbb{E}[X_k] = 1$).

Exemple V.25. Comment calculer $m = \int_0^1 g(x)\,dx$, où g est mesurable bornée, à l'aide d'un ordinateur ? Le générateur de nombres aléatoires sur un ordinateur fournit $(U_n(\omega), n \in \mathbb{N}^*)$, une réalisation d'une suite de variables aléatoires que l'on considère indépendantes et de loi uniforme sur $[0,1]$. Ensuite, par abus, on admet que l'on a bien convergence de la moyenne empirique $\frac{1}{n}\sum_{k=1}^{n} g(U_k(\omega))$ vers la moyenne $\mathbb{E}[g(U_1)]$ pour cette réalisation particulière ω. On a donc pour n assez grand :

$$\frac{1}{n}\sum_{k=1}^{n} g(U_k(\omega)) \simeq \mathbb{E}[g(U)] = \int_0^1 g(x)\,dx = m.$$

Cette méthode, dite méthode de Monte Carlo (1949), est particulièrement efficace si on considère des intégrales sur $[0,1]^d$ avec d grand, comparativement à d'autres méthodes numériques. Le but du paragraphe V.6 est de préciser cette approximation. ◊

Démonstration. **Loi forte des grands nombres**. On donne une démonstration sous des hypothèses plus fortes de la loi forte des grands nombres : on suppose que $\mathbb{E}[X_k^4] < \infty$. On pose $Y_k = X_k - \mathbb{E}[X_k]$ et $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$. Les v.a. Y_k sont indépendantes et identiquement distribuées. On a $\mathbb{E}[Y_k^4] < \infty$, et $\mathbb{E}[Y_k] = 0$. Un simple calcul montre que :

$$\begin{aligned}\mathbb{E}\left[\bar{Y}_n^4\right] &= \frac{1}{n^4}\,\mathbb{E}\left[\left(\sum_{k=1}^{n} Y_k\right)^4\right] \\ &= \frac{1}{n^4}\left(3n(n-1)\mathbb{E}\left[Y_1^2\right]^2 + n\mathbb{E}\left[Y_1^4\right]\right) \\ &\leq \frac{3\mathbb{E}[Y_1^4]}{n^2}.\end{aligned}$$

Comme $\mathbb{E}[Y_1^4] < \infty$, on en déduit par convergence monotone que $\mathbb{E}\left[\sum_{n=1}^{\infty} \bar{Y}_n^4\right] < \infty$. Cela implique que p.s. :

$$\sum_{n=1}^{\infty} \bar{Y}_n^4 < \infty.$$

En particulier, presque sûrement, les termes de la série tendent vers 0, cela signifie que p.s. $\lim_{n\to\infty} \bar{Y}_n = 0$. Comme $\lim_{n\to\infty} \mathbb{E}[\bar{Y}_n^4] = 0$, on a également, grâce à l'inégalité de Jensen que $\lim_{n\to\infty} \mathbb{E}[|\bar{Y}_n|] = 0$. Cela démontre la dernière partie du théorème de la loi forte des grands nombres quand $\mathbb{E}[X_k^4] < \infty$. □

Que se passe-t-il si les v.a. sont toujours indépendantes et de même loi mais pas intégrables ? On distingue suivant les trois raisons pour lesquelles une v.a. réelle

n'est pas intégrable : sa partie positive est intégrable mais pas sa partie négative ; sa partie négative est intégrable mais pas sa partie positive ; ni sa partie positive ni sa partie négative ne sont intégrables. La loi de Cauchy appartient à ce dernier cas ; voir le comportement de la moyenne empirique de v.a. de Cauchy i.i.d. sur la figure V.3 (on peut montrer que la loi de Cauchy vérifie la condition 3-c de la proposition V.26). On admet le résultat suivant.

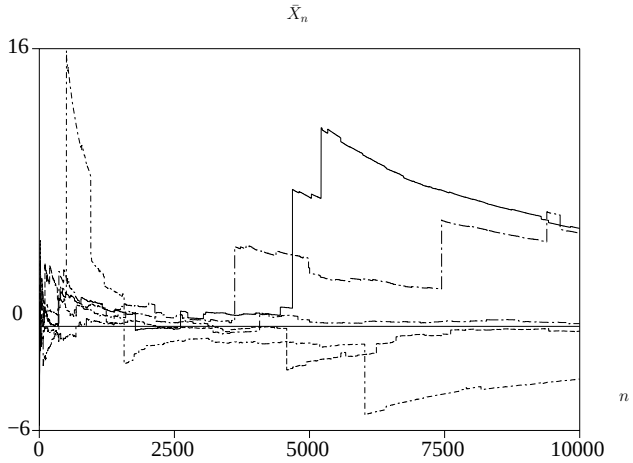


Figure V.3. Plusieurs réalisations de la suite $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, où les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes de loi de Cauchy de paramètre $a = 1$.

Proposition V.26. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes et de même loi. On suppose que $\mathbb{E}[|X_1|] = +\infty$. On note $X_1^+ = X_1 \mathbf{1}_{\{X_1>0\}}$ la partie positive de X_1 et $X_1^- = |X_1|\,\mathbf{1}_{\{X_1<0\}}$ la partie négative.*

1. *Si $\mathbb{E}[X_1^+] < \infty$, alors p.s. $\lim_{n\to\infty} \bar{X}_n = -\infty$.*
2. *Si $\mathbb{E}[X_1^-] < \infty$, alors p.s. $\lim_{n\to\infty} \bar{X}_n = +\infty$.*
3. *Si $\mathbb{E}[X_1^-] = +\infty$ et $\mathbb{E}[X_1^+] = +\infty$, alors une des trois assertions suivantes est vraie*[1] *:*
   a) p.s. $\lim_{n\to\infty} \bar{X}_n = +\infty$;
   b) p.s. $\lim_{n\to\infty} \bar{X}_n = -\infty$;
   c) p.s. $\liminf_{n\to\infty} \bar{X}_n = -\infty$ et $\limsup_{n\to\infty} \bar{X}_n = +\infty$.

1. Voir la note 6.8.19 p.231 de la monographie *Limit Theorems of Probability Theory* de V. Petrov. Oxford University Press, 1995.

V.5 Estimations de lois

On observe une réalisation, $X_1, \ldots, X_n$, de n variables aléatoires indépendantes et de même loi. Dans ce qui suit, on donne des méthodes pour obtenir des renseignements sur la loi de X_1.

V.5.1 Variables aléatoires discrètes

On suppose pour simplifier que X_1 est à valeurs dans $\mathbb{N}$. Il s'agit d'un exemple générique pour les v.a.d. Pour tout $i \in \mathbb{N}$, on a d'après la LFGN que la **fréquence empirique** :

$$\hat{p}_i(n) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k = i\}}$$

converge p.s. vers la fréquence $\mathbb{E}[\mathbf{1}_{\{X_1=i\}}] = \mathbb{P}(X_1 = i) = p_i$. Par σ-additivité on peut permuter les opérateurs $\forall i$ et p.s. On en déduit que p.s. pour tout $i \in \mathbb{N}$:

$$\lim_{n\to\infty} \hat{p}_i(n) = p_i.$$

Ainsi la loi empirique $(\hat{p}_i(n), i \in \mathbb{N})$ converge p.s. vers la loi de X_1 : $(p_i, i \in \mathbb{N})$. Ceci est illustré sur la figure V.4. On peut également montrer que p.s. :

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} |\hat{p}_i(n) - p_i| = 0.$$

V.5.2 Variables aléatoires réelles

La loi de X_1 est caractérisée par sa fonction de répartition F. On définit la **fonction de répartition empirique**, F_n, pour $x \in \mathbb{R}$ par :

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \le x\}} = \frac{1}{n} \text{ Card } \{k \in \{1, \ldots, n\}; X_k \le x\}.$$

Pour tout $x \in \mathbb{R}$, on déduit de la LFGN que la suite $(F_n(x), n \in \mathbb{N})$ converge p.s. vers $F(x) = \mathbb{P}(X \le x)$. La convergence de la fonction de répartition empirique F_n vers la fonction de répartition (cf. le théorème V.27 ci-dessous) est illustrée par la simulation de la figure V.5. On admet le résultat suivant plus général.

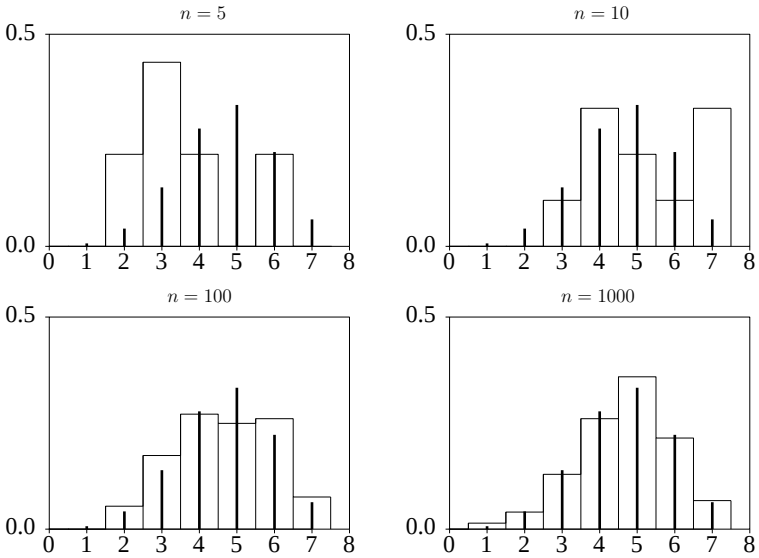


Figure V.4. Comparaison de la loi binomiale $\mathcal{B}(7, 2/3)$ (bâtons) ($p_k = \mathbb{P}(X_i = k), k \in \mathbb{N}$) et de la loi empirique (pavés) ($\hat{p}_k(n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i=k\}}, k \in \mathbb{N}$) d'une réalisation de n variables aléatoires indépendantes et de même loi $\mathcal{B}(7, 2/3)$.

Théorème V.27 (Glivenko-Cantelli)**.** *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires réelles indépendantes de même loi et de fonction de répartition F. On a p.s. :*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0,$$

où F_n est la fonction de répartition empirique de l'échantillon $X_1, \ldots, X_n$.

V.5.3 Variables aléatoires à densité

Peut-on estimer la densité f de la loi de de la v.a. X_1 à valeurs dans $\mathbb{R}^d$? Il n'existe pas de méthode directe comme dans les deux cas précédents.

Estimation à noyau.

Soit une fonction K, appelée noyau, symétrique positive bornée d'intégrale 1. On choisit souvent l'un des noyaux suivants :

– noyau uniforme sur $[-1/2, 1/2]^d$: $K(u) = \mathbf{1}_{[-1/2,1/2]^d}(u)$,

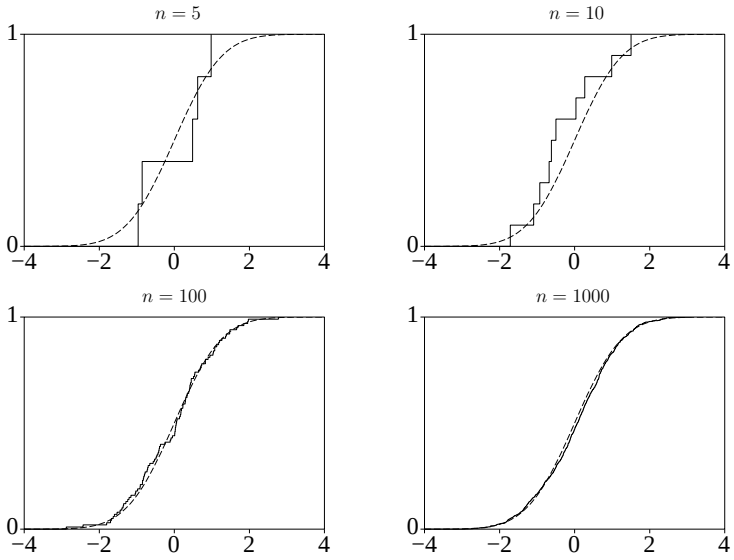


Figure V.5. Comparaison de la fonction de répartition exacte (pointillés) et d'une réalisation de la fonction de répartition empirique (trait plein) de $(X_1, \ldots, X_n)$, variables aléatoires indépendantes et de même loi $\mathcal{N}(0,1)$.

– noyau d'Epanechnikov : $K(u) = c_d(1-|u|^2)\mathbf{1}_{\{|u|\leq 1\}}$ où $c_d = \dfrac{d+2}{2\int_{\mathbb{R}^d} \mathbf{1}_{\{|x|\leq 1\}}\, dx}$,

– noyau gaussien : $K(u) = (2\pi)^{-d/2} \exp\left(-|u|^2/2\right)$.

Pour $x \in \mathbb{R}^d$, on considère l'estimation [2] à noyau de $f(x)$ par :

$$f_n(x) = \frac{1}{nh^d} \sum_{k=1}^{n} K\left(\frac{x - X_k}{h}\right), \tag{V.1}$$

où le paramètre $h > 0$ est appelé la largeur de fenêtre (ou de bande). À x fixé, $f_n(x)$ converge p.s. vers $\frac{1}{h^d}\int_{\mathbb{R}^d} K\left(\frac{x-y}{h}\right) f(y)dy$. Si h est petit et si la densité f est continue, alors cela donne une bonne approximation de $f(x)$. On admet le résultat de convergence uniforme suivant.

Théorème V.28. *On suppose que la densité f est uniformément continue sur $\mathbb{R}^d$. Soit $(h_n, n \in \mathbb{N})$ une suite de réels positifs telle que $\lim_{n\to\infty} h_n = 0$ et $\lim_{n\to\infty} nh_n^d = +\infty$, alors p.s. on a :*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} |f_n(x) - f(x)| = 0,$$

où f_n est l'estimation à noyau (V.1) de la densité f, avec pour paramètre $h = h_n$.

2. Voir par exemple M.P. Wand et M.C. Jones : *Kernel Smoothing*, Chapman & Hall, 1995.

Le choix du noyau est empiriquement moins crucial que le choix de la largeur de bande h_n, qui dépend de n et éventuellement de la réalisation $X_1, \ldots, X_n$. En particulier si h est petit, l'approximation f_n est très irrégulière, et si h est grand l'approximation f_n est loin de f, ainsi on ne distingue plus les modes (maxima) de f (cf. la figure V.6).

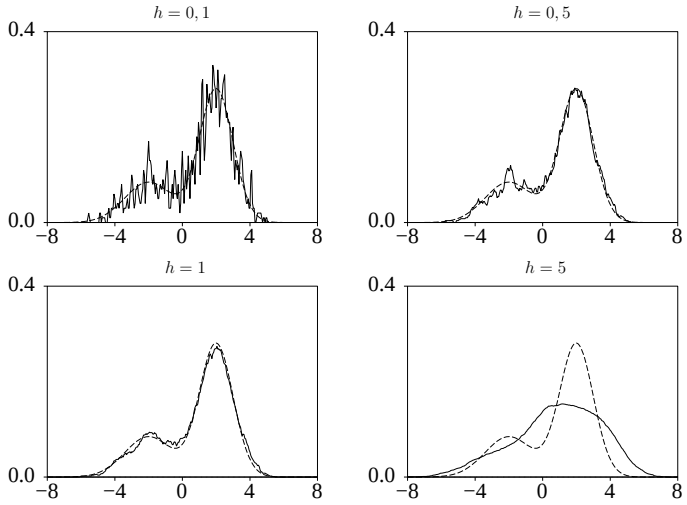


Figure V.6. Comparaison, en fonction de la largeur h de la fenêtre, de la densité (pointillés) d'une variable aléatoire réelle et de l'estimation par noyau (trait plein), avec le noyau uniforme, d'une réalisation de $n = 1000$ v.a. indépendantes et de même loi.

Estimation par histogramme.

Pour simplifier, on suppose que les v.a. continues sont réelles. On considère alors la partition de $\mathbb{R}$ en intervalles $[ih, (i+1)h[$, où $i \in \mathbb{Z}$. On approche les fréquences :

$$p_i^h = \mathbb{P}(ih \le X_1 < (i+1)h) = \int_{[ih,(i+1)h[} f(x)\, dx,$$

par les fréquences empiriques $\hat{p}_i^h(n) = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} \mathbf{1}_{\{ih \le X_k < (i+1)h\}}$. Pour h petit, p_i^h est une bonne approximation de $hf(ih)$ si f est continue. Et pour n grand on a par la loi forte des grands nombres (voir aussi le paragraphe ci-dessus concernant les v.a.d.) que p.s. pour tout $i \in \mathbb{Z}$, $\lim_{n\to\infty} \hat{p}_i^h(n) = p_i^h$. L'estimation de f, par l'histogramme empirique $(\hat{p}_i^h(n), i \in \mathbb{Z})$ converge pour $n \to \infty$, $h \to 0$ et $nh \to \infty$,

au moins si f est continue. La figure V.7 en donne une illustration. Le choix de h est à nouveau crucial. Cette méthode contrairement à l'estimation à noyau ne tient pas compte pour l'estimation de $f(ih)$ des point voisins de l'intervalle $[ih, (i+1)h[$. On peut éventuellement considérer des intervalles $[x_i, x_i + h_i[$ dont les longueurs h_i sont variables et sont fonction de $X_1, \ldots, X_n$.

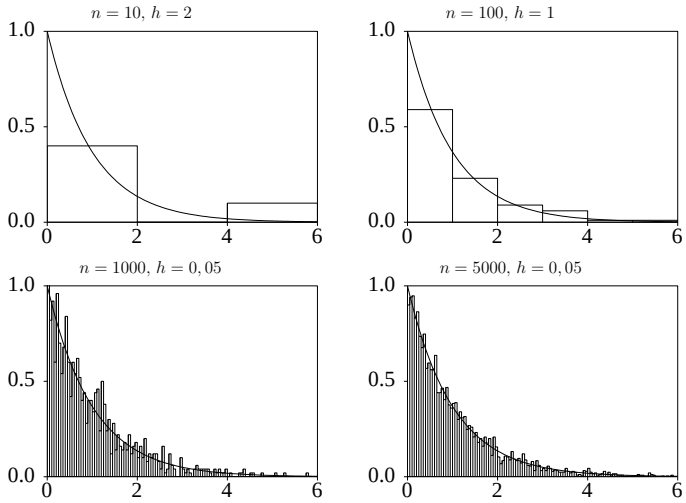


Figure V.7. Comparaison de la densité exponentielle $\mathcal{E}(1)$ et de l'histogramme empirique d'une réalisation de n variables aléatoires indépendantes et de même loi exponentielle $\mathcal{E}(1)$.

V.6 Théorème central limite

Le théorème central limite (TCL) précise la vitesse de convergence de la loi forte des grands nombres. Il s'agit d'un des résultats fondamentaux de la théorie des probabilités.

Théorème V.29 (Théorème central limite)**.** *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de v.a. réelles* ***i.i.d.*** *(indépendantes et de même loi). On suppose qu'elles sont de* ***carré intégrable*** *(*$\mathbb{E}[X_n^2] < \infty$*). On pose* $\mu = \mathbb{E}[X_n]$, $\sigma^2 = \mathrm{Var}(X_n)$ *et la moyenne empirique* $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$*. La suite de v.a.* $\sqrt{n}(\bar{X}_n - \mu) = \dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}}$ *converge en loi vers la loi gaussienne* $\mathcal{N}(0, \sigma^2)$ *:*

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}} \quad \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} \quad \mathcal{N}(0, \sigma^2).}$$

Comme les v.a. X_n sont i.i.d. et intégrables, on sait que la moyenne empirique $\bar{X}_n$ converge p.s. vers $\mu = \mathbb{E}[X_n]$. Le théorème central limite donne la vitesse de convergence. La figure V.8 permet d'observer visuellement la convergence de la densité de $Y_n = \sqrt{n}(\bar{X}_n - \mu)$ vers la densité gaussienne si les v.a. $(X_n, n \in \mathbb{N}^*)$ sont continues. De manière plus générale, on peut observer la convergence de l'histogramme empirique de Y_n vers la densité gaussienne, voir la figure V.9.

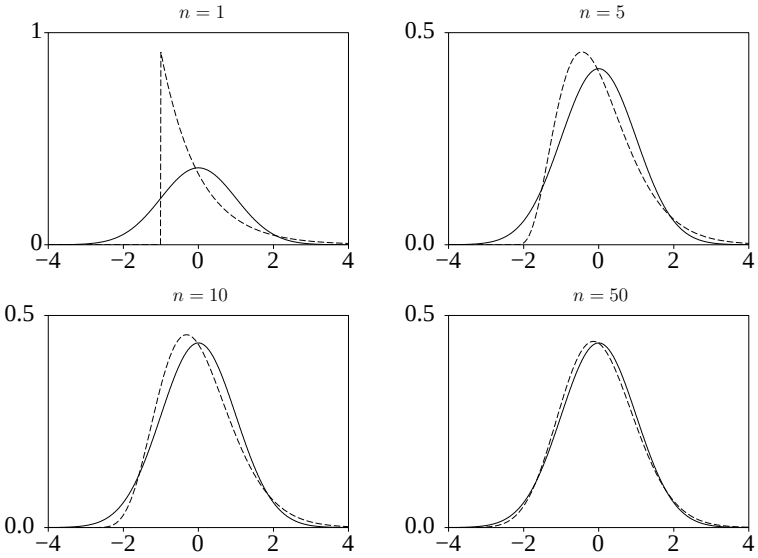


Figure V.8. Comparaison de la densité de la loi gaussienne $\mathcal{N}(0,1)$ (en trait plein) et de la densité (pointillés) de $Y_n = \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}[X_1]\right)$, où les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes de loi exponentielle de paramètre $\lambda = 1$ (on a $\mathbb{E}[X_1] = 1/\lambda = 1$ et $\mathrm{Var}(X_1) = 1/\lambda^2 = 1$).

Comme la loi gaussienne est une loi à densité, elle ne charge pas les points de discontinuité de la fonction indicatrice $\mathbf{1}_{[-a\sigma,a\sigma]}$. Après avoir remarqué que :

$$\mathbb{E}\left[\mathbf{1}_{[-a\sigma,a\sigma]}(\sqrt{n}[\bar{X}_n - \mu])\right] = \mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right),$$

on déduit du théorème V.29 et de la proposition V.19 le corollaire suivant.

Corollaire V.30. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes, identiquement distribuées et de carré intégrable. On pose $\mu = \mathbb{E}[X_n]$ et $\sigma^2 = \mathrm{Var}(X_n)$. Alors, pour $a > 0$, on a :*

$$\boxed{\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) \underset{n\to\infty}{\longrightarrow} \int_{-a}^{a} \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}.}$$

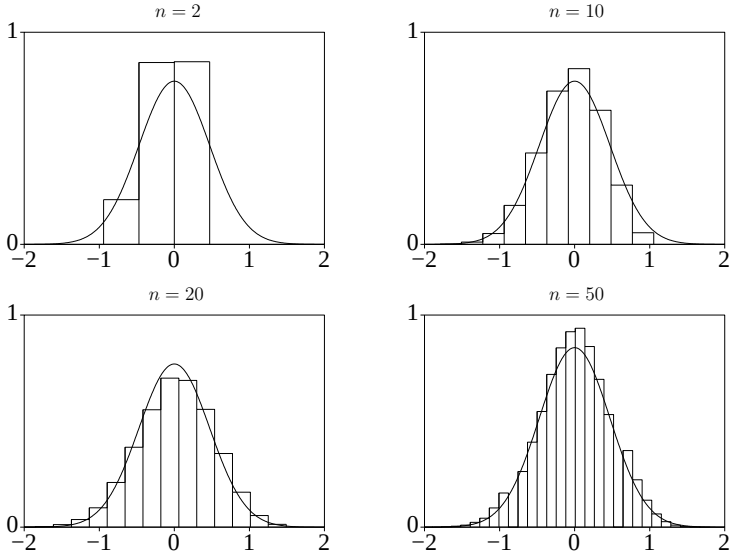


Figure V.9. Comparaison de la densité de $\mathcal{N}(0, p(1-p))$ et de l'histogramme empirique ($m = 10\,000$ réalisations) de $Y_n = \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}[X_1]\right)$, où les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes de loi de Bernoulli de paramètre $p = 2/3$ (on a $\mathbb{E}[X_1] = p = 2/3$ et $\mathrm{Var}(X_1) = p(1-p) = 2/9$).

Si l'on désire donner une approximation de μ, à l'aide de la moyenne empirique $\bar{X}_n$, on peut fournir un intervalle aléatoire :

$$I_n = \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]$$

qui contient la valeur de la moyenne, μ, avec une probabilité asymptotique $1-\alpha = \mathbb{P}(|Y| \leq a)$, où Y est de loi $\mathcal{N}(0,1)$. L'intervalle I_n est appelé **intervalle de confiance** de μ de **niveau asymptotique** $1-\alpha$. Les valeurs les plus couramment utilisées sont $\alpha = 5\%$, soit $1-\alpha = 95\%$ et $a \simeq 1.96$, et $\alpha = 1\%$, soit $1-\alpha = 99\%$ et $a \simeq 2.58$.

Les hypothèses concernant le TCL peuvent être affaiblies. On peut affaiblir par exemple l'hypothèse d'indépendance ou d'égalité en loi (voir l'exercice V.4). Chaque jeu d'hypothèses fournit un nouveau théorème central limite. La littérature est vaste dans ce domaine.

Exercice V.4.
En s'inspirant de la démonstration du théorème central limite, montrer le résultat suivant. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. réelles indépendantes telles que $\mathbb{E}[X_n] =$

0 et $\mathbb{E}[|X_n|^3] < \infty$ pour tout $n \in \mathbb{N}$. On note $\sigma_n^2 = \operatorname{Var} X_n = \mathbb{E}[X_n^2]$. On suppose que $\sum_{n=1}^{\infty} \sigma_n^2 = +\infty$ et que $\lim_{n\to\infty} \frac{\sum_{k=1}^n \mathbb{E}[|X_n|^3]}{(\sum_{k=1}^n \sigma_k^2)^{3/2}} = 0$. Montrer que :

$$\frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \quad \xrightarrow[n\to\infty]{\textbf{en loi}} \quad \mathcal{N}(0,1).$$

△

Démonstration. **Théorème central limite**. On pose $Y_k = X_k - \mu$. En particulier, $\mathbb{E}[Y_k] = 0$ et $\mathbb{E}[Y_k^2] = \sigma^2$. Les v.a. réelles $(Y_k, k \in \mathbb{N}^*)$ sont i.i.d. En utilisant les propriétés des fonctions caractéristiques, on a :

$$\begin{aligned}\psi_{\sqrt{n}(\bar{X}_n - \mu)}(u) &= \psi_{\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k}(u) \\ &= \left[\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)\right]^n.\end{aligned}$$

On rappelle que pour $x \in \mathbb{R}$:

$$\left|\mathrm{e}^{ix} - 1 - ix + \frac{x^2}{2}\right| \leq \min\left(\frac{|x|^3}{6}, x^2\right).$$

Donc on a l'égalité $\mathrm{e}^{ity} = 1 + ity - \frac{t^2y^2}{2} + h(y)$, où $|h(y)|$ est majorée par $t^2 \min\left(\frac{t|y|^3}{6}, y^2\right)$. En particulier, on a avec $t = u/\sqrt{n}$ et $y = Y_1$:

$$\mathrm{e}^{i\frac{uY_1}{\sqrt{n}}} = 1 + i\frac{u}{\sqrt{n}}Y_1 - \frac{u^2Y_1^2}{2n} + h_n(Y_1)$$

avec $|h_n(Y_1)| \leq \frac{u^2}{n} \min\left(\frac{u|Y_1|^3}{6\sqrt{n}}, Y_1^2\right)$. On remarque que $h_n(Y_1)$ est intégrable car $|h_n(Y_1)| \leq \frac{u^2}{n}Y_1^2$. En prenant l'espérance dans l'égalité ci-dessus, il vient :

$$\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{u^2\sigma^2}{2n} + \mathbb{E}[h_n(Y_1)].$$

Comme la v.a. $n|h_n(Y_1)|$ est uniformément majorée par $u^2Y_1^2$, qui est intégrable, et que $\lim_{n\to\infty} n|h_n(Y_1)| \leq \lim_{n\to\infty} \frac{u^3Y_1^3}{6\sqrt{n}} = 0$ p.s., par le théorème de convergence dominée, on a :

$$\lim_{n\to\infty} n\mathbb{E}[h_n(Y_1)] = 0. \tag{V.2}$$

Le terme $\mathbb{E}[h_n(Y_1)]$ est *a priori* complexe. Pour étudier la suite de terme général $\left(1-\frac{u^2\sigma^2}{2n}+\mathbb{E}[h_n(Y_1)]\right)^n$ on a recours au lemme suivant qui se démontre facilement par récurrence.

Lemme V.31. *Soit $(a_k, k\in\mathbb{N}^*)$ et $(b_k, k\in\mathbb{N}^*)$ des suites de nombres complexes de modules inférieurs à 1 ($|a_k|\le 1$ et $|b_k|\le 1$ pour tout $k\in\mathbb{N}^*$). On a :*

$$\left|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\right| \le \sum_{k=1}^n |a_k - b_k|\,.$$

Pour n assez grand tel que $u^2\sigma^2/n<1$, on a :

$$\begin{aligned}\left|\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)^n - \left(1-\frac{u^2\sigma^2}{2n}\right)^n\right| &\le \sum_{k=1}^n \left|\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right) - 1 + \frac{u^2\sigma^2}{2n}\right| \\ &= n\,|\mathbb{E}[h_n(Y_1)]|\,.\end{aligned}$$

Il vient :

$$\begin{aligned}\left|\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)^n - \mathrm{e}^{-\frac{u^2\sigma^2}{2}}\right| &\le \left|\psi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)^n - \left(1-\frac{u^2\sigma^2}{2n}\right)^n\right| \\ &\quad + \left|\left(1-\frac{u^2\sigma^2}{2n}\right)^n - \mathrm{e}^{-\frac{u^2\sigma^2}{2}}\right| \\ &\le n\,|\mathbb{E}[h_n(Y_1)]| + \left|\left(1-\frac{u^2\sigma^2}{2n}\right)^n - \mathrm{e}^{-\frac{u^2\sigma^2}{2}}\right|.\end{aligned}$$

On remarque que $\left(1-\frac{u^2\sigma^2}{2n}\right)^n = \mathrm{e}^{n\log\left(1-\frac{u^2\sigma^2}{2n}\right)}$ converge vers $\mathrm{e}^{-\frac{u^2\sigma^2}{2}}$ quand n tend vers l'infini. On déduit donc de (V.2) que : pour tout $u\in\mathbb{R}$,

$$\lim_{n\to\infty} \psi_{\sqrt{n}(\bar{X}_n-\mu)}(u) = \mathrm{e}^{-\frac{u^2\sigma^2}{2}} = \psi_{\mathcal{N}(0,\sigma^2)}(u).$$

On a donc la convergence en loi de $\sqrt{n}(\bar{X}_n-\mu)$ vers la loi gaussienne $\mathcal{N}(0,\sigma^2)$. □

V.7 Autour du théorème central limite (I)

On énonce le théorème de Slutsky, puis on donne une application importante concernant la méthode de Monte-Carlo.

Théorème V.32 (Slutsky)**.** *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires qui converge en loi vers une variable aléatoire X. Soit $(Y_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires définies sur le même espace que les v.a. X_n, qui converge presque sûrement ou en probabilité ou en loi vers la* ***constante*** *a. Alors la suite $((X_n, Y_n), n \in \mathbb{N}^*)$ converge en loi vers le couple (X, a).*

Démonstration. Comme la convergence presque sûre et la convergence en probabilité impliquent la convergence en loi, on suppose simplement que la suite $(Y_n, n \in \mathbb{N})$ converge en loi vers la constante a. On utilise les fonctions caractéristiques. On veut donc montrer que $\psi_{X_n,Y_n}(u,v) = \mathbb{E}\left[\mathrm{e}^{i(u,X_n)+i(v,Y_n)}\right]$ converge vers $\mathrm{e}^{i(v,a)}\,\mathbb{E}\left[\mathrm{e}^{i(u,X_n)}\right] = \psi_{X,a}(u,v)$ pour tout couple (u,v). On a :

$$\begin{aligned}
&|\psi_{X_n,Y_n}(u,v) - \psi_{X,a}(u,v)| \\
&\qquad = \left|\mathbb{E}\left[\mathrm{e}^{i(u,X_n)}\left(\mathrm{e}^{i(v,Y_n)} - \mathrm{e}^{i(v,a)}\right) + \left(\mathrm{e}^{i(u,X_n)} - \mathrm{e}^{i(u,X)}\right)\mathrm{e}^{i(v,a)}\right]\right| \\
&\qquad \le \mathbb{E}\left[\left|\mathrm{e}^{i(u,X_n)}\left(\mathrm{e}^{i(v,Y_n)} - \mathrm{e}^{i(v,a)}\right)\right|\right] + \left|\mathrm{e}^{i(v,a)}\,\mathbb{E}\left[\mathrm{e}^{i(u,X_n)} - \mathrm{e}^{i(u,X)}\right]\right| \\
&\qquad = \mathbb{E}\left[\left|\mathrm{e}^{i(v,Y_n)} - \mathrm{e}^{i(v,a)}\right|\right] + |\psi_{X_n}(u) - \psi_X(u)|\,.
\end{aligned}$$

Comme la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers X, la quantité $|\psi_{X_n}(u) - \psi_X(u)|$ converge vers 0, quand $n \to \infty$. Par ailleurs, la fonction $g(y) = \left|\mathrm{e}^{i(v,y)} - \mathrm{e}^{i(v,a)}\right|$ est une fonction continue bornée. Comme la suite $(Y_n, n \in \mathbb{N}^*)$ converge en loi vers a, on en déduit que $\mathbb{E}[g(Y_n)]$ converge vers $\mathbb{E}[g(a)] = 0$ quand $n \to \infty$. Ceci termine la démonstration. □

Exercice V.5.
Soit $(X_n, n \in \mathbb{N})$ une suite de matrices aléatoires de taille $k \times p$ qui converge en loi vers X. Soit $(Y_n, n \in \mathbb{N})$ une suite de matrices aléatoires de taille $l \times k$ qui converge en loi vers une matrice constante Σ. Montrer que la suite de matrices aléatoires $(Y_n X_n, n \in \mathbb{N})$ converge en loi vers la matrice aléatoire ΣX. △

Correction V.5. La suite $((X_n, Y_n), n \in \mathbb{N})$ converge en loi vers (X, Σ) d'après le théorème de Slutsky. Enfin l'application φ qui, à deux matrices, x de taille $k \times p$ et y de taille $l \times k$, associe $\varphi(x,y) = yx$ est continue. On déduit du corollaire V.21 que la suite $(\varphi(X_n, Y_n), n \in \mathbb{N})$ converge en loi vers $\varphi(X, \Sigma)$. ▲

Application. Si on désire donner une approximation de $\mu = \mathbb{E}[X]$, où X est une v.a. réelle de carré intégrable que l'on sait simuler, on peut utiliser la moyenne

empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, où les v.a. $(X_i, i \in \mathbb{N}^*)$ sont i.i.d. de même loi que X. Par la loi forte des grands nombres, on a p.s. $\bar{X}_n \simeq \mu$. On peut de plus, grâce au théorème central limite fournir un intervalle de confiance de niveau asymptotique $1-\alpha$: $\left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]$, où la constante a est déterminée par $\mathbb{P}(|Z| \leq a) = 1-\alpha$, Z étant une v.a. gaussienne de loi $\mathcal{N}(0,1)$. La constante a est le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. En général, si on ne connaît pas l'espérance de X, il est rare que l'on connaisse σ^2 la variance de X. Il faut donc remplacer σ dans l'intervalle de confiance par une estimation. Comme $\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, on déduit de la loi forte des grands nombres que la suite $(\sigma_n^2, n \in \mathbb{N}^*)$, définie par :

$$\boxed{\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2,}$$

converge p.s. vers σ^2.

On déduit du théorème de Slutsky que $((\sqrt{n}(\bar{X}_n - \mu), \sigma_n^2), n \in \mathbb{N}^*)$ converge en loi vers $(\sigma Z, \sigma^2)$. Enfin, la fonction $f(x,y) = x/\sqrt{|y|}$ si $y \neq 0$ et $f(x,y) = 0$ sinon, admet $\mathbb{R} \times \{0\}$ comme ensemble de points de discontinuités. Si $\sigma^2 > 0$, on a $\mathbb{P}((\sigma Z, \sigma^2) \in \mathbb{R} \times \{0\}) = 0$. On déduit donc du corollaire V.21, que la suite $(f(\sqrt{n}(\bar{X}_n - \mu), \sigma_n^2), n \in \mathbb{N}^*)$ converge en loi vers Z qui a pour loi $\mathcal{N}(0,1)$. En particulier, une nouvelle utilisation de la proposition V.19, avec la fonction $h(r) = \mathbf{1}_{[-a,a]}(r)$, assure que si $\sigma \neq 0$ alors :

$$\boxed{\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]\right) \underset{n\to\infty}{\longrightarrow} \int_{-a}^{a} \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}.}$$

Pour n grand, μ est dans l'intervalle de confiance aléatoire $\left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]$ avec une probabilité proche de $1-\alpha = \mathbb{P}(|Z| \leq a)$. La quantité $1-\alpha$ est le **niveau asymptotique** de l'intervalle de confiance. On choisit souvent $1-\alpha = 95\%$ avec $a \simeq 1.96$ ou $1-\alpha = 99\%$ avec $a \simeq 2.58$. Dans la figure V.10, on trace l'évolution de la moyenne empirique et de l'intervalle de confiance associé en fonction du nombre de données. On remarque que si $\sigma = 0$, alors les v.a. X_n sont p.s. égales à μ. On a alors $\sigma_n = 0$ p.s. et $\mathbb{P}(\bar{X}_n = \mu) = 1$. ◊

On retiendra que **toute estimation** à l'aide d'observations ou de simulations (méthode de Monte-Carlo) doit être fournie avec un **intervalle de confiance**.

On peut enfin se poser la question de la validité de l'intervalle asymptotique : on a $\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]\right) \simeq \mathbb{P}(|Z| \leq a)$, mais quelle est la précision

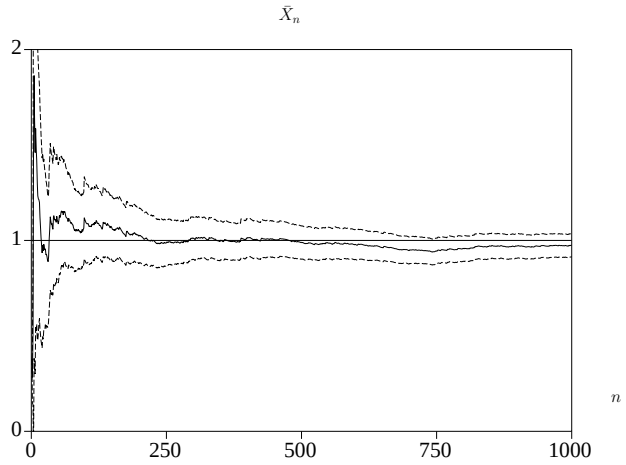


Figure V.10. Intervalle de confiance : Réalisations de la suite $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, où les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes de loi exponentielle de paramètre $\lambda = 1$ (on a alors $\mathbb{E}[X_k] = 1$) ainsi que de l'intervalle de confiance $I_n = \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]$ de niveau asymptotique 95% ($a \simeq 1.96$), où la variance empirique est définie par $\sigma_n^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}_n^2$.

de cette égalité asymptotique ? Il s'agit de résultats dus à Berry et Esséen dans les années 1941-1942 pour l'intervalle de confiance construit avec σ et de résultats récents (1996) pour l'intervalle de confiance construit avec σ_n, l'estimation de σ.

On se contente d'énoncer les deux résultats suivants.

Théorème V.33. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes et identiquement distribuées. On suppose $\mathbb{E}[|X_1|^3] < \infty$. On note $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1) > 0$ et $\mu_3 = \mathbb{E}[|X_1 - \mu|^3]$. Alors pour tout $a \in \mathbb{R}$, $n \geq 1$, on a :*

$$\boxed{\left| \mathbb{P}\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq a \right) - \int_{-\infty}^{a} \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}} \right| \leq \frac{C\mu_3}{\sigma^3 \sqrt{n}},}$$

où la constante C est universelle (i.e. indépendante de a et de la loi de X_1) avec $(2\pi)^{-1/2} \leq C < 0,8$.

Pour a grand, on peut remplacer la constante C par une fonction de a, $C(a)$, telle que $\lim_{|a|\to\infty} C(a) = 0$.

Enfin si on remplace la variance σ par son estimation σ_n :

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2,$$

on a le résultat suivant.

Théorème V.34. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes et identiquement distribuées. On suppose $\mathbb{E}[|X_1|^3] < \infty$. On note $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1) > 0$ et $\mu_3 = \mathbb{E}[|X_1 - \mu|^3]$. Alors pour tout $a \in \mathbb{R}$, $n \geq 1$, on a :*

$$\boxed{\left|\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma_n} \leq a\right) - \int_{-\infty}^a \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}\right| \leq \frac{C'\mu_3}{\sigma^3\sqrt{n}},}$$

où la constante C' est universelle (i.e. indépendante de a et de la loi de X_1).

En particulier si μ_3/σ^3 (ou une approximation de μ_3/σ^3) est élevée, cela suggère que la convergence du théorème central limite peut être mauvaise.

Toutefois comme les majorations des théorèmes V.33 et V.34 sont indépendantes de la loi des v.a. (la constante C est universelle), elles sont dans bien des cas très grossières. Elles donnent cependant le bon ordre pour des v.a. de Bernoulli.

Exemple. On considère des v.a. X_n i.i.d. de loi de Bernoulli de paramètre $p \in]0,1[$. La vitesse de convergence du théorème de Berry-Esséen est donc $\frac{C[(1-p)^2+p^2]}{\sqrt{(1-p)p}\sqrt{n}} = C_p n^{-1/2}$. La constante C_p est grande si le paramètre p est proche de 1 ou de 0. Dans ces deux cas les intervalles de confiance sont de mauvaise qualité. Leurs niveaux exacts sont très différents du niveau asymptotique. ◊

V.8 Autour du théorème central limite (II)

On pourra omettre ce paragraphe dans une première lecture.

On décrit dans un premier temps le comportement asymptotique de certains évènements rares.

Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles, indépendantes, de même loi, intégrables et telles que $\mathbb{E}[X_n] = 0$. On note $\bar{X}_n$ la moyenne empirique. On déduit de la loi forte des grands nombres que $\lim_{n\to\infty} \mathbb{P}(\bar{X}_n \geq a) = 0$ si $a > 0$. Si de plus les v.a. sont de carré intégrable avec $\sigma^2 = \mathrm{Var}(X_1) = \mathbb{E}[X_1^2] > 0$, alors on connaît la limite de $\mathbb{P}(\bar{X}_n \geq a/\sqrt{n})$. Cette limite est une conséquence du TCL :

$$\mathbb{P}(\bar{X}_n \geq a/\sqrt{n}) \underset{n\to\infty}{\longrightarrow} \int_{a/\sigma}^{\infty} \mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}}.$$

Si on remplace $\sqrt{n}$ par n^α, avec $\alpha > 1/2$, on en déduit que $\lim_{n\to\infty} \mathbb{P}(\bar{X}_n \geq a/n^\alpha) = 1/2$. En revanche si $\alpha \in [0, 1/2[$, alors la limite ci-dessus est nulle. En fait on peut préciser, dans ce cas, à quelle vitesse $\mathbb{P}(\bar{X}_n \geq a/n^\alpha)$ converge vers 0. Il s'agit d'une vitesse exponentielle. Pour $\alpha \in]0, 1/2[$, on parle de **déviations modérées**, et pour $\alpha = 0$, on parle de **grandes déviations**.

Théorème V.35 (Déviations modérées)**.** *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles, indépendantes, de même loi. On suppose qu'il existe $c > 0$ tel que $\mathbb{E}\left[\mathrm{e}^{c|X_1|}\right] < \infty$ et que $\mathbb{E}[X_1] = 0$. On note $\sigma^2 = \mathrm{Var}(X_1)$. Pour tout $a > 0$, $\alpha \in]0, 1/2[$, on a :*

$$\lim_{n\to\infty} n^{2\alpha-1} \log \mathbb{P}(\bar{X}_n \geq a/n^\alpha) = -\frac{a^2}{2\sigma^2}.$$

Cela revient formellement à utiliser le TCL, avec $an^{\frac{1}{2}-\alpha}$ au lieu de a, puis à donner un équivalent logarithmique de $\displaystyle\int_{an^{\frac{1}{2}-\alpha}/\sigma}^{\infty} \mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}}$. Un phénomène différent apparaît pour $\alpha = 0$.

Théorème V.36 (Grandes déviations)**.** *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles, indépendantes, de même loi. Pour tout $a > 0$, on a :*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq a) = -I(a),$$

où la fonction I est définie par $I(x) = \sup\left\{zx - \log \mathbb{E}\left[\mathrm{e}^{zX}\right] ; z \in \mathbb{R}\right\}$.

L'étude des déviations s'est fortement développée depuis les années 1980.

Les deux théorèmes ci-dessus ainsi que le théorème central limite sont évident dans le cas où les v.a. sont gaussiennes.
Exemple. Pour la loi exponentielle de paramètre λ, on a $I(a) = \lambda a - \log(\lambda a + 1)$. Pour la loi gaussienne $\mathcal{N}(0, \sigma^2)$, on a $I(a) = \dfrac{a^2}{2\sigma^2}$. ◊

Une deuxième approche consiste, dans le cas où les v.a. i.i.d. $(X_n, n \in \mathbb{N}^*)$ sont continues, à regarder f_n, la densité de la moyenne empirique, et de faire un développement limité en n de f_n ou de la fonction de répartition $\int_{-\infty}^{x} f_n(z)\, dz$. On

parle de développement d'Edgeworth. Cette méthode imaginée à la fin du XIX^{e} siècle, a connu un fort développement dans les années 1970-1980.

Théorème V.37. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. continues indépendantes et de même loi. On suppose que $\mathbb{E}\left[|X_1|^{k+2}\right] < \infty$ et, pour des raisons de normalisation, que $\mathbb{E}[X_1] = 0$, et $\mathrm{Var}(X_1) = 1$. Il existe des polynômes $p_1, \dots, p_k$, tels que : pour tout $x \in \mathbb{R}$,*

$$\mathbb{P}\left(\sqrt{n}\bar{X}_n \le x\right) = \int_{-\infty}^{x} \mathrm{e}^{-z^2/2} \ \frac{dz}{\sqrt{2\pi}} + n^{-1/2} p_1(x)\,\mathrm{e}^{-x^2/2} + \cdots + n^{-k/2} p_k(x)\,\mathrm{e}^{-x^2/2} + o\left(n^{-k/2}\right).$$

On pose $\mu_j = \mathbb{E}[X^j]$. On peut montrer que le polynôme p_j est de degré $3j-1$ et ne dépend que de $\mu_3, \dots, \mu_{j+2}$ et qu'en particulier :

$$p_1(x) = -\frac{1}{6}\,\mu_3(x^2-1) \text{ et } p_2(x) = -x\left(\frac{1}{24}\,\mu_4(x^2-1) + \frac{1}{72}\,\mu_3^2(x^4-10x^2+15)\right).$$

V.9 Résumé

- On a les implications suivantes concernant les convergences :

$$\text{CV p.s.} \implies \text{CV en probabilité} \implies \text{CV en loi.}$$

- Soit h une fonction mesurable bornée. On note C l'ensemble des points de discontinuité de h. Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. qui converge en loi vers X. Si $\mathbb{P}(X \in C) = 0$, alors on a :

$$\lim_{n\to\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)].$$

- La suite de v.a. réelles ou vectorielles $(X_n, n \in \mathbb{N}^*)$ **converge en loi** vers X si et seulement si l'une des trois conditions suivantes est satisfaite
  - Pour toute fonction g **continue bornée**, on a :

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)].$$

  - Pour tout u, on a : $\lim_{n\to\infty} \psi_{X_n}(u) = \psi_X(u)$.
  - Les fonctions de répartition de X_n, F_n, convergent vers F, la fonction de répartition de X, sauf peut-être aux points de discontinuité de F.

– Théorème de Slutsky. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires qui converge en loi vers une variable aléatoire X. Soit $(Y_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires définies sur le même espace que les v.a. X_n, qui converge presque sûrement ou en probabilité ou en loi vers la **constante** a. Alors la suite $((X_n, Y_n), n \in \mathbb{N}^*)$ converge en loi vers le couple (X, a).

– **Loi forte des grands nombres**. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles ou vectorielles **indépendantes, de même loi et intégrables** ($\mathbb{E}[|X_1|] < \infty$). On pose $\mu = \mathbb{E}[X_n]$. Alors on a :

$$\boxed{\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\text{p.s.}} \mu.}$$

De plus on a $\mathbb{E}[\left|\bar{X}_n - \mu\right|] \to 0$ quand $n \to \infty$.

– **Théorème central limite**. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles **indépendantes, de même loi et de carré intégrable** ($\mathbb{E}[X_n^2] < \infty$). On pose $\mu = \mathbb{E}[X_n]$, $\sigma^2 = \mathrm{Var}(X_n)$ et la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. On a :

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{en loi}} \mathcal{N}(0, \sigma^2).}$$

– **Intervalle de confiance** : Sous les hypothèses du théorème central limite, on note :

$$\boxed{\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}$$

une estimation de σ^2 (σ_n^2 converge p.s. vers σ^2 par la loi forte des grands nombres). On considère l'intervalle de confiance :

$$\boxed{I_n = \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right].}$$

Alors, si $\sigma > 0$, on a :

$$\boxed{\mathbb{P}(\mu \in I_n) \underset{n\to\infty}{\longrightarrow} \int_{-a}^{a} \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}}.}$$

On peut remplacer σ_n par la vraie valeur σ dans I_n sans changer le résultat. Le niveau asymptotique de l'intervalle de confiance est donné par $\displaystyle\int_{-a}^{a} \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}}$.

– **Toute estimation à l'aide d'observations ou de simulations** (méthode de Monte-Carlo) **doit être fournie avec un intervalle de confiance.**

– Théorème de Berry-Esséen. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes et identiquement distribuées. On suppose $\mathbb{E}[|X_1|^3] < \infty$. On note $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1) > 0$ et $\mu_3 = \mathbb{E}[|X_1 - \mu|^3]$. Alors pour tout $a \in \mathbb{R}$, $n \geq 1$, on a :

$$\left| \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq a \right) - \int_{-\infty}^{a} \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}} \right| \leq \frac{C\mu_3}{\sigma^3\sqrt{n}},$$

où la constante $C < 0,8$ est universelle. Il existe un résultat similaire quand on remplace σ par σ_n.

V.10 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice V.1 p. 132, *Exercice* V.2 p. 134, *Exercice* V.3 p. 136, *Exercice* V.4 p. 147, *Exercice* V.5 p. 150.

Exercice V.6.
Soit $(X_i, i \in \mathbb{N}^*)$, une suite de variables aléatoires indépendantes de loi exponentielle de paramètre λ.

1. Calculer la loi de $Z_n = \min\{X_i; i \leq n\}$. (On regardera $\mathbb{P}(Z_n \geq z)$.) En déduire que la suite $(nZ_n, n \geq 1)$ est constante en loi et donc converge en loi.
2. Calculer la loi de $Y_n = \max\{X_i; i \leq n\}$, puis la loi de $\tilde{Y}_n = \lambda Y_n - \log n$. Montrer que pour $n \geq 2$, $(n-1)\log(1 - n^{-1}\,\mathrm{e}^{-x}) \leq -\frac{1}{2}\,\mathrm{e}^{-x}$ si $x > -\log n$. Montrer, en utilisant la définition de la convergence en loi, que la suite $(\tilde{Y}_n, n \geq 1)$ converge en loi. Vérifier que la fonction de répartition de la loi limite est $F(x) = \mathrm{e}^{-\,\mathrm{e}^{-x}}$, pour $x \in \mathbb{R}$. La loi limite est appelée loi de Gumbel.

△

Exercice V.7.
Calculer la limite en loi de la suite $(X_n, n \in \mathbb{N}^*)$, où :

1. X_n suit une loi binomiale de paramètre $(n, \lambda/n)$ $(\lambda > 0, n > \lambda)$.
2. $X_n = T_n/n$, où T_n suit une loi géométrique de paramètre λ/n. Traiter ensuite le cas où T_n suit une loi de k-ième succès pour un schéma de Bernoulli de paramètre λ/n.
3. $X_n = (S_n - pn)/\sqrt{n}$, où S_n suit une loi binomiale de paramètre (n, p), avec $p \in]0, 1[$.

△

Exercice V.8.
Soit $(X_i, i \in \mathbb{N}^*)$, une suite de variables aléatoires i.i.d de loi $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. On désire étudier les convergences en loi de :

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad Y_n = \sum_{k=1}^{n} \frac{1}{2^k}\, X_k \quad \text{et} \quad Z_n = \frac{1}{n}\sum_{k=1}^{n} \sqrt{k} X_k.$$

1. Rappeler la convergence de $(\bar{X}_n, n \in \mathbb{N}^*)$.

2. Montrer que Y_n converge p.s. vers une variable aléatoire Y. Montrer par récurrence que :
$$\sin\left(\frac{\lambda}{2^n}\right)\prod_{k=1}^{n}\cos\left(\frac{\lambda}{2^k}\right)=\frac{\sin(\lambda)}{2^n}.$$
En déduire que Y_n converge en loi vers la loi uniforme sur $[-1,1]$. En déduire la loi de Y.
3. Montrer par récurrence que si $(a_k, k\in\mathbb{N}^*)$ et $(b_k, k\in\mathbb{N}^*)$ sont des suites de complexes tels que pour tout $k\in\mathbb{N}^*$, $|a_k|\leq 1$ et $|b_k|\leq 1$, alors :
$$\left|\prod_{k=1}^{n}a_k-\prod_{k=1}^{n}b_k\right|\leq\sum_{k=1}^{n}|a_k-b_k|.$$
Montrer qu'il existe M tel que $\forall x\in\mathbb{R}$, $\left|\cos(x)-\exp\left(-x^2/2\right)\right|\leq Mx^4$. En déduire que : pour tout $u\in\mathbb{R}$,
$$\lim_{n\to\infty}\left|\psi_n(u)-\prod_{k=1}^{n}\mathrm{e}^{-\frac{ku^2}{2n^2}}\right|=0,$$
où ψ_n est la fonction caractéristique de Z_n. Montrer que Z_n converge en loi. Donner la loi limite.

△

Exercice V.9.
Soit $(X_n, n\in\mathbb{N}^*)$ une suite de variables aléatoires réelles continues, indépendantes et de même loi. On suppose que la densité, f, de leur loi est bornée, symétrique, continue en 0 et telle que $f(0)>0$.

1. Montrer que la suite de variable aléatoire $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{X_i}$ converge en loi vers une variable aléatoire de Cauchy dont on déterminera le paramètre en fonction de $f(0)$. On rappelle que :
$$\int_0^\infty\frac{1-\cos(u)}{u^2}\,du=\lim_{T\to\infty}\int_0^T\frac{\sin(u)}{u}\,du=\frac{\pi}{2}.$$
2. En déduire que la moyenne harmonique empirique $\tilde{X}_n=\dfrac{n}{\sum_{i=1}^{n}\frac{1}{X_i}}$ converge en loi vers une loi de Cauchy.

△

Exercice V.10.
Soit $(X_n, n \in \mathbb{N}^*)$, une suite de variables aléatoires de loi exponentielle qui converge en loi. Déterminer la loi de la limite. △

Exercice V.11.
Soit $(X_n, n \in \mathbb{N}^*)$, une suite de variables aléatoires indépendantes de loi de Cauchy de paramètre a.

1. Rappeler la fonction caractéristique de X_1. Calculer la loi de $M_n = \frac{1}{n}\sum_{i=1}^n X_i$. La suite $(M_n, n \in \mathbb{N}^*)$ est-elle convergente ?
2. Calculer la loi de $M_{n+p} - M_p$. En déduire le comportement de $\mathbb{P}(|M_{2n} - M_n| > 1)$ quand $n \to \infty$?
3. Montrer par l'absurde que la suite $(M_n, n \in \mathbb{N}^*)$ ne converge pas en probabilité.

△

Exercice V.12.
Soit X_N une variable aléatoire de loi hypergéométrique de paramètre (N, m, n). On rappelle que X_N représente le nombre de boules blanches obtenues lors d'un tirage sans remise de n boules hors d'une urne contenant m boules blanches et $N - m$ boules noires.

1. Vérifier que $\mathbb{P}(X_N = k) = \dfrac{C_m^k C_{N-m}^{n-k}}{C_N^n} = \dfrac{C_n^k C_{N-n}^{m-k}}{C_N^m}$ pour $n - N + m \le k \le m$ et $n \ge k \ge 0$.
2. On suppose que le nombre de boules blanches, m, est fixé, n et N tendent vers $+\infty$ avec $\lim_{N\to+\infty} n/N = p \in [0,1]$ (p est la proportion limite du nombre de boules obtenues lors du tirage). Montrer que la suite $(X_N, N \in \mathbb{N}^*)$ converge en loi vers la loi binomiale de paramètre (m, p).
3. On suppose que le tirage est de taille n est fixé, m et N tendent vers $+\infty$ avec $\lim_{N\to+\infty} m/N = \theta \in [0,1]$ (θ est la proportion limite du nombre de boules blanches dans l'urne). Montrer que la suite $(X_N, N \in \mathbb{N}^*)$ converge en loi vers la loi binomiale de paramètre (n, θ).

△

Exercice V.13.
Soit $(X_m, m \in \mathbb{N}^*)$ une suite de variables aléatoires discrètes à valeurs dans $\mathbb{N}$. On pose $p_k(m) = \mathbb{P}(X_m = k)$. On suppose que $\lim_{m\to\infty} p_k(m) = p_k$.

1. Montrer que si la suite $(X_m, m \in \mathbb{N}^*)$ converge en loi alors $\sum_{k=0}^\infty p_k = 1$.
2. On souhaite démontrer la réciproque. On suppose que $\sum_{k=0}^\infty p_k = 1$. Montrer que pour tous $N \in \mathbb{N}$ et $\varepsilon > 0$, il existe m_0 tel que pour tout $m \ge m_0$, $\sum_{k>N} p_k(m) < \varepsilon$. Montrer que la suite $(X_m, m \in \mathbb{N}^*)$ converge en loi.

△

Exercice V.14.
Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. continues. On note f_n la densité de la loi de X_n. On suppose que la suite $(f_n, n \in \mathbb{N}^*)$ est dominée par une fonction h intégrable. On suppose de plus que cette suite converge vers une limite f. Montrer que la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi et identifier la limite. (La réciproque est fausse en général.) △

Exercice V.15.
Soit $(Y_m, m \in \mathbb{N}^*)$ une suite de v.a. de loi binomiale de paramètre (m, p_m). On suppose que $m \to \infty$ et $\lim_{m\to\infty} p_m m = \theta \in]0, \infty[$. Montrer que la suite $(Y_m, m \in \mathbb{N}^*)$ converge en loi vers la loi de Poisson de paramètre θ. △

Exercice V.16.
Soit $(Z_m, m \in \mathbb{N}^*)$ une suite de v.a. de Poisson de paramètre θ_n. On suppose que l'on a $\lim_{n\to\infty} \theta_n/n = \alpha \in]0, 1[$. Montrer que la suite de v.a. $(Z_n - \theta_n)/\sqrt{n}$ converge en loi vers une limite que l'on identifiera. △

Exercice V.17.
Soit $(X_i, i \in \mathbb{N}^*)$ une suite de v.a. indépendantes de loi exponentielle de paramètre λ. On pose $S_n = (\sum_{i=1}^n X_i - n\lambda^{-1})/\sqrt{n}$.

1. Montrer que S_n converge en loi vers $\mathcal{N}(0, 1/\lambda^2)$ quand $n \to \infty$.
2. Calculer $\mathbb{P}(S_n \in]a, b[)$ et montrer que :

$$\mathbb{P}\left(S_n \in]a, b[\right) = \int_a^b \frac{1}{\sqrt{2\pi}\lambda} \, \mathrm{e}^{-y^2/2\lambda^2} \, dy + \frac{1}{\sqrt{n}} \int_a^b \left(y - \frac{y^3}{3}\right) \frac{1}{\sqrt{2\pi}\lambda} \, \mathrm{e}^{-y^2/2\lambda^2} \, dy + O\left(\frac{1}{n}\right).$$

Comparer ce résultat avec le théorème de Berry-Esséen ou le développement d'Edgeworth.

△

Exercice V.18.
On compte en 1996 environ 3,9 millions de naissances aux USA qui se décomposent en 1 990 480 garçons et 1 901 014 filles.

1. Donner une estimation de la probabilité qu'un nouveau-né soit un garçon. Quels sont les chiffres significatifs ?
2. Dire s'il est raisonnable de penser qu'il naît autant de filles que de garçons ?

△

Exercice V.19.
Précision des sondages.

1. À quelle précision peut prétendre un sondage sur deux candidats effectué sur un échantillon de 1 000 personnes ? Est-ce que ce résultat dépend de la taille de la population ?
2. En Floride, pour l'élection présidentielle américaine 2000, on compte 6 millions de votants. Sachant qu'il y a eu environ 4 000 voix d'écart, quel est le nombre de personnes qu'il aurait fallu interroger dans un sondage pour savoir avec 95% de chance qui allait être le vainqueur ?

△

Exercice V.20.
On effectue n séries de 400 tirages de pile ou face avec une pièce équilibrée. On observe les fréquences empiriques de pile $F_1, \ldots, F_n$ dans ces séries.

1. Quelle est (approximativement) la loi de probabilité du nombre N de ces fréquences $(F_i, 1 \leq i \leq n)$ qui ne vérifient pas la condition $0.45 < F_i < 0.55$, lorsque $n = 20$?
2. Est-il plus probable que $N = 0$, que $N = 1$ ou que $N \geq 2$?

△

Exercice V.21.
Lors d'une opération de calcul, l'ordinateur renvoie un résultat arrondi à la précision ε de la machine près. On effectue n opérations et on désire connaître la précision du résultat. À l'opération i, on commet une erreur E_i. On suppose que les erreurs sont aléatoires indépendantes et de même loi uniforme. On désire étudier l'erreur finale $S_n = \sum_{i=1}^{n} E_i$ en fonction de la technique d'arrondi choisie.

1. Les résultats sont tronqués (arrondis vers le bas). Quel est le support de la loi de E_i ? Quel est l'asymptotique de S_n pour n grand ? On désire garantir une précision finale de 10^{-12}, avec $\varepsilon = 10^{-16}$. Quel est l'ordre de grandeur du nombre d'opérations autorisé ?
2. Les résultats sont arrondis au nombre machine le plus proche (arrondis vers le bas ou vers le haut). Quel est le support de la loi de E_i ? Quelle est l'asymptotique de S_n pour n grand ? On désire garantir, avec une probabilité supérieure à $99,9\%$, une précision finale de 10^{-12}, avec $\varepsilon = 10^{-16}$. Quel est l'ordre de grandeur du nombre d'opérations autorisé ?

△

Exercice V.22.
Soit X et X' deux variables aléatoires réelles, indépendantes, de carré intégrable et de même loi $\mathcal{L}$. On désire déterminer les lois $\mathcal{L}$ vérifiant la condition (C) suivante :

$$\frac{X+X'}{\sqrt{2}} \quad \text{et } X \text{ ont même loi.}$$

1. Calculer $\mathbb{E}[X]$.
2. Vérifier que la loi gaussienne centrée réduite vérifie la condition (C).
3. En utilisant le théorème de la limite centrale, déterminer les seules lois vérifiant la condition (C). (Indication : déterminer par récurrence la loi de $2^{n/2}\sum_{k=1}^{2^n} X_k$.)

△

Exercice V.23.
Le temps d'attente chez le médecin suit une loi uniforme sur $[0,\theta]$, où θ est un paramètre inconnu propre à chaque médecin. Un nouveau médecin s'installe dans votre ville et vous désirez estimer la valeur de θ qui lui correspond. Vous interrogez ses patients sur leur temps d'attente. On modélise les temps d'attente par des variables aléatoires $(X_i, i \geq 1)$ indépendantes de loi uniforme sur $[0,\theta]$.

1. Calculer $\mathbb{E}[X_i]$. En déduire une méthode pour estimer θ. Quelle est votre précision ?
2. On pose $M_n = \max_{i\in\{1,\dots,n\}}(X_i)$. Établir la convergence p.s de la suite $(M_n, n \geq 1)$. Montrer qu'elle converge en probabilité vers θ. Conclusion.
3. Montrer en utilisant les fonctions de répartitions que $(n(\theta - M_n), n \geq 1)$ converge en loi. Déterminer sa limite. En déduire un intervalle de confiance pour θ.
4. Comparer les deux intervalles de confiance pour un même niveau asymptotique.

△

Exercice V.24.
Soit $(X_n, n \geq 1)$ une suite de variables aléatoires, à valeurs dans $\{1,\dots,k\}$, indépendantes de même loi définie par $\mathbb{P}(X_n = i) = p_i$ pour $i \in \{1,\dots,k\}$ avec $\sum_{i=1}^k p_i = 1$. On note $p_n(i_1,\dots,i_n) = \mathbb{P}(X_1 = i_1,\dots,X_n = i_n)$ la probabilité d'observer la séquence $i_1,\dots,i_n$. Montrer que presque sûrement la suite $(-\frac{1}{n}\log p_n(X_1,\dots,X_n), n \geq 1)$ converge vers l'entropie de la loi de X_n définie par $H = -\sum_{i=1}^k p_i \log p_i$. △

Exercice V.25.
Pour calculer $m = \int_0^1 g(x)dx$ on utilise souvent des méthodes de simulation appelées méthodes de Monte-Carlo. Le but de cet exercice est de comparer plusieurs

méthodes. Soit g une fonction mesurable à valeurs dans $[0,1]$ et intégrable. Soit X, Y des variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On pose $U = \mathbf{1}_{\{Y \leq g(X)\}}$, $V = g(X)$ et $W = 2^{-1}[g(X) + g(1-X)]$.

1. Calculer l'espérance et la variance de U,V et W.
2. Comparer les comportements asymptotiques de :

$$A_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq g(X_i)\}},$$
$$B_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i),$$
$$C_n = \frac{1}{2n}\sum_{i=1}^{n} [g(X_i) + g(1-X_i)],$$

   où les variables aléatoires $(X_i, Y_i; i \geq 1)$ sont indépendantes et de même loi uniforme sur $[0,1]$.
3. Donner un intervalle de confiance pour les estimations de m à l'aide de A_n et à l'aide de B_n. Quelle méthode choisissez vous ?

On suppose dorénavant que g est croissante.

4. Montrer que $[g(X)-g(Y)][g(1-X)-g(1-Y)] \leq 0$. En déduire que $\mathbb{E}[g(X)g(1-X)] \leq \mathbb{E}[g(X)]^2$.
5. En pratique, il est numériquement coûteux d'évaluer la fonction g. En particulier il faut comparer l'estimation de m à partir de B_{2n} et l'estimation de m à partir de C_n qui toutes deux utilisent $2n$ évaluations de la fonction g. Quelle méthode choisissez vous ?

△

VI

Vecteurs gaussiens

Les vecteurs gaussiens sont une extension au cas vectoriel des variables aléatoires gaussiennes réelles, voir le paragraphe VI.1 pour leur définition et leurs propriétés. En particulier, au paragraphe VI.3, on peut énoncer le théorème central limite dans un cadre vectoriel. Dans le paragraphe VI.2, on calcule des lois qui sont naturellement associées aux vecteurs gaussiens et qui, avec les vecteurs gaussiens, sont également très utilisées en statistiques, voir par exemple les modèles de régression linéaire du paragraphe IX.7.

VI.1 Définition et propriétés

Soit X une variable aléatoire de loi gaussienne $\mathcal{N}(m, \sigma^2)$, avec $m \in \mathbb{R}$ et $\sigma > 0$. Sa fonction caractéristique est $\psi_X(u) = \mathrm{e}^{ium - \frac{\sigma^2 u^2}{2}}$. Si σ tend vers 0, alors la fonction caractéristique converge vers $\psi(u) = \mathrm{e}^{ium}$. On reconnaît la fonction caractéristique de la v.a. constante égale à m. Par **convention**, on dira que c'est également une v.a. de loi gaussienne $\mathcal{N}(m, 0)$. Dans ce cas on parle de v.a. gaussienne dégénérée.

La proposition suivante assure que la famille des lois gaussiennes est stable pour la convergence en loi.

Proposition VI.1. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. gaussiennes de loi $\mathcal{N}(m_n, \sigma_n^2)$. La suite converge en loi si et seulement si $m_n \underset{n\to\infty}{\longrightarrow} m \in \mathbb{R}$ et $\sigma_n \underset{n\to\infty}{\longrightarrow} \sigma \in [0, \infty[$. Et la loi limite est la loi gaussienne $\mathcal{N}(m, \sigma^2)$.*

La démonstration qui suit peut être omise dans une première lecture.

Démonstration. On suppose que la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi. La suite de fonctions caractéristiques $\psi_n(u) = \mathrm{e}^{ium_n - \frac{\sigma_n^2 u^2}{2}}$ converge quand n tend vers l'infini, pour tout $u \in \mathbb{R}$. La partie radiale de ψ_n converge donc. Cela entraîne que σ_n converge. On note $\sigma \geq 0$ la limite. Comme la limite de la suite ψ_n est continue en 0, la limite σ est finie. Pour la partie imaginaire, on suppose dans un premier temps que la suite $(m_n, n \in \mathbb{N}^*)$ est bornée. En prenant la limite inférieure $\underline{m}$ et la limite supérieure $\overline{m}$, on obtient que pour tout $u \in \mathbb{R}$, $\mathrm{e}^{iu\underline{m}} = \mathrm{e}^{iu\overline{m}}$, ce qui implique que $\underline{m} = \overline{m} = m$ et que la suite $(m_n, n \in \mathbb{N}^*)$ converge vers m. Si la suite $(m_n, n \in \mathbb{N}^*)$ est non bornée, il existe une sous-suite $(m_{n_k}, k \in \mathbb{N}^*)$, qui diverge vers $+\infty$ (ou $-\infty$). Soit f une fonction à valeurs réelles, de classe C^1 et à support compact, par exemple dans $[-A, A]$. À l'aide d'une intégration par partie, il vient :

$$\int_{\mathbb{R}} \mathrm{e}^{im_{n_k}u} f(u)\, du = -\frac{1}{im_{n_k}} \int_{\mathbb{R}} \mathrm{e}^{im_{n_k}u} f'(u)\, du.$$

Donc, on a $\left|\int_{\mathbb{R}} \mathrm{e}^{im_{n_k}u} f(u)\, du\right| \leq 2A \left\|f'\right\|_\infty / m_{n_k}$ et $\lim_{k\to\infty} \int_{\mathbb{R}} \mathrm{e}^{im_{n_k}u} f(u)\, du = 0$. La convergence en loi implique que $\lim_{n\to\infty} \mathrm{e}^{im_n u} = g(u)$ et g est continue. Par convergence dominée, on en déduit que pour toute fonction f, de classe C^1, à support compact, $\int_{\mathbb{R}} f(u)g(u)\, du = 0$. En choisissant une suite de fonctions de classe C^1, à support compact, dominée par 1 qui converge vers $\overline{g(u)}\mathbf{1}_{[a,b](u)}$, on déduit du théorème de convergence dominée que $\int_a^b |g(u)|^2\, du = 0$ pour tout $a, b \in \mathbb{R}$. On en déduit que $g = 0$ presque partout. Or $|g(u)| = 1$. Ce résultat est donc absurde. On ne peut pas extraire de sous-suite divergente.

On a donc obtenu que $\lim_{n\to\infty} m_n = m \in \mathbb{R}$, et $\lim_{n\to\infty} \sigma_n = \sigma \in [0, \infty[$. En particulier, on a :

$$\lim_{n\to\infty} \psi_n(u) = \mathrm{e}^{ium - \frac{\sigma^2 u^2}{2}} = \psi_{\mathcal{N}(m,\sigma^2)}(u).$$

La loi limite est une loi gaussienne. □

On utilise les notations suivantes. Soit $d \geq 1$. On note $\langle \cdot, \cdot \rangle$ le produit scalaire et $|\cdot|$ la norme sur $\mathbb{R}^d$. Pour une matrice $M = (M_{k,j}; 1 \leq k \leq d, 1 \leq j \leq p)$ de taille $d \times p$, on note M^t sa transposée et pour $v = (v_1, \ldots, v_p) \in \mathbb{R}^p$, $Mv = ((Mv)_1, \ldots, (Mv)_d)$ désigne le vecteur de $\mathbb{R}^d$ tel que $(Mv)_k = \sum_{j=1}^p M_{k,j} v_j$. Ainsi pour $u \in \mathbb{R}^d$, on a $\langle u, Mv \rangle = \langle M^t u, v \rangle$.

Définition VI.2. *Un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est un* ***vecteur gaussien*** *si toute combinaison linéaire de ses coordonnées est une v.a. réelle gaussienne :*

$$\boxed{\forall a \in \mathbb{R}^d, \quad \textit{la loi de } \langle a, X \rangle \textit{ est une loi gaussienne.}}$$

En particulier, en choisissant $a = e_i$, où le vecteur e_i a toutes ses coordonnées nulles sauf la i-ième qui est égale à 1, on obtient que si $X = (X_1, \ldots, X_d)$ est un vecteur gaussien, alors $X_i = \langle X, e_i \rangle$ est une variable aléatoire gaussienne.

On remarque que tout vecteur gaussien de dimension 1 est une v.a. réelle gaussienne éventuellement dégénérée. On rappelle que si $\mathcal{L}(X) = \mathcal{N}(m, \sigma^2)$, alors $\mathcal{L}(aX) = \mathcal{N}(am, a^2\sigma^2)$.

Exemple VI.3. Soit $X_1, \ldots, X_d$ des v.a. gaussiennes **indépendantes**. On suppose que la loi de X_k est la loi gaussienne $\mathcal{N}(m_k, \sigma_k^2)$. Alors le vecteur $X = (X_1, \ldots, X_d)$ est un vecteur gaussien. En effet, soit $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$. On calcule la fonction caractéristique de $\langle a, X \rangle$:

$$\begin{aligned}\psi_{\langle a,X\rangle}(u) &= \mathbb{E}\left[\mathrm{e}^{iu\sum_{k=1}^d a_k X_k}\right] \\ &= \prod_{k=1}^d \mathbb{E}\left[\mathrm{e}^{iua_kX_k}\right] \quad \text{par indépendance} \\ &= \prod_{k=1}^d \mathrm{e}^{iua_km_k - \frac{a_k^2\sigma_k^2u^2}{2}} = \exp\left(iu\langle a, m\rangle - \frac{\langle a, \Lambda a\rangle u^2}{2}\right),\end{aligned}$$

où $m = (m_1, \ldots, m_d)$ et $\Lambda = \mathrm{Diag}\,(\sigma_1^2, \ldots, \sigma_d^2)$ est une matrice diagonale. On en déduit que la loi de $\langle a, X \rangle$ est la loi gaussienne $\mathcal{N}(\langle a, m\rangle, \langle a, \Lambda a\rangle)$. Donc X est un vecteur gaussien. ◊

Définition VI.4. *La* ***matrice de covariance*** *de deux v.a. vectorielles de carré intégrable, X à valeurs dans $\mathbb{R}^d$ et Y à valeurs dans $\mathbb{R}^n$, est la matrice* $\mathrm{Cov}(X, Y) = V = (V_{k,l}, 1 \le k \le d, 1 \le l \le n)$ *définie par :*

$$\boxed{V_{k,l} = \mathbb{E}[X_kY_l] - \mathbb{E}[X_k]\mathbb{E}[Y_l] = \mathrm{Cov}(X_k, Y_l)}.$$

Remarque. Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien. Le vecteur X est de carré intégrable. En effet, on a $|X|^2 = \sum_{i=1}^d X_i^2$, où les v.a. X_i sont de carré intégrable car de loi gaussienne. Par linéarité, on en déduit que $\mathbb{E}[|X|^2] < \infty$. ◊

On peut caractériser la loi des vecteurs gaussiens de manière simple.

Proposition VI.5. *Le vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est un vecteur gaussien si et seulement si il existe un vecteur $\mu \in \mathbb{R}^d$ et une matrice V de taille $d \times d$, symétrique positive ($V^t = V$ et $\langle x, Vx\rangle \ge 0, \quad \forall x \in \mathbb{R}^d$) tels que :*

$$\boxed{\psi_X(u) = \mathrm{e}^{i\langle \mu, u\rangle - \frac{\langle u, Vu\rangle}{2}}, \quad \forall u \in \mathbb{R}^d.}$$

De plus le vecteur X est de carré intégrable et on a $\mu = \mathbb{E}[X]$ et $V = \mathrm{Cov}(X, X)$. Enfin, pour tout $a \in \mathbb{R}^d$, la loi de la v.a. réelle $\langle a, X\rangle$ est la loi gaussienne $\mathcal{N}(\langle a, \mu\rangle, \langle a, Va\rangle)$.

Si X est un vecteur gaussien de moyenne μ et de matrice de covariance V, alors on note sa loi $\mathcal{N}(\mu, V)$.

Démonstration. On a vu que X_i est une variable aléatoire gaussienne. En particulier elle est de carré intégrable. Ceci implique que X est de carré intégrable. On remarque que pour $u \in \mathbb{R}^d$:

$$\psi_X(u) = \mathbb{E}[\mathrm{e}^{i\langle u, X\rangle}] = \psi_{\langle u, X\rangle}(1).$$

Or par définition, $\langle u, X\rangle$ est de loi gaussienne. On calcule les paramètres de cette loi :

$$\mathbb{E}[\langle u, X\rangle] = \mathbb{E}\left[\sum_{k=1}^{d} u_k X_k\right] = \sum_{k=1}^{d} u_k \mathbb{E}[X_k] = \langle u, \mu\rangle, \quad \text{où} \quad \mu = \mathbb{E}[X],$$

et par bilinéarité :

$$\begin{aligned}\mathrm{Var}(\langle u, X\rangle) = \mathrm{Var}\left(\sum_{k=1}^{d} u_k X_k\right) &= \sum_{\substack{1\le k\le d\\ 1\le l\le d}} u_k u_l \left(\mathbb{E}[X_k X_l] - \mathbb{E}[X_k]\mathbb{E}[X_l]\right)\\ &= \langle u, \mathrm{Cov}(X, X) u\rangle.\end{aligned}$$

On en déduit, en posant $V = \mathrm{Cov}(X, X)$, que :

$$\psi_{\langle u, X\rangle}(1) = \mathrm{e}^{i\langle u, \mu\rangle - \frac{\langle u, Vu\rangle}{2}}.$$

Il reste à vérifier que la matrice V est symétrique (évident d'après sa construction) et positive. On remarque que $\langle u, Vu\rangle = \mathrm{Var}(\langle u, X\rangle)$, et cette quantité est toujours positive.

La démonstration de la réciproque est similaire à la démonstration de l'exemple VI.3. □

La proposition suivante permet de caractériser facilement l'indépendance pour les composantes d'un vecteur gaussien.

Proposition VI.6. *Soit (X,Y) un **vecteur gaussien**. Alors, on a :*

$$\boxed{\text{les v.a. } X \text{ et } Y \text{ sont indépendantes} \iff \mathrm{Cov}(X,Y)=0.}$$

L'hypothèse "(X,Y) est un vecteur gaussien" est cruciale. Voir les exercices II.7 et VI.1 pour des contre-exemples.

Démonstration. On suppose que $\mathrm{Cov}(X,Y)=0$. En utilisant la fonction caractéristique du vecteur (X,Y), on obtient :

$$\psi_{(X,Y)}(u,v) = \exp\left(i\langle u,\mu_X\rangle + i\langle v,\mu_Y\rangle - \frac{\langle u,V_X,u\rangle + \langle v,V_Y v\rangle + 2\langle u,V_{X,Y}v\rangle}{2}\right),$$

où $\mu_X=\mathbb{E}[X]$, $\mu_Y=\mathbb{E}[Y]$, $V_X=\mathrm{Cov}(X,X)$, $V_Y=\mathrm{Cov}(Y,Y)$ et $V_{X,Y}=\mathrm{Cov}(X,Y)$. Comme $V_{X,Y}=0$, on remarque que :

$$\psi_{(X,Y)}(u,v)=\psi_X(u)\psi_Y(v), \quad \forall u,\forall v.$$

Grâce au 2 de la proposition IV.6 (pour des v.a. vectorielles), cela implique que les v.a. X et Y sont indépendantes.

La réciproque est vraie car X et Y sont de carré intégrable. □

Exercice VI.1.
Soit X une variable aléatoire réelle de loi $\mathcal{N}(0,1)$. Soit ε une variable aléatoire discrète indépendante de X et telle que $\mathbb{P}(\varepsilon=1)=\mathbb{P}(\varepsilon=-1)=1/2$. On pose $Y=\varepsilon X$. Déterminer la loi de Y. Calculer $\mathrm{Cov}(X,Y)$. Calculer $\mathbb{E}[X^2Y^2]$ et $\mathbb{E}[X^2]\mathbb{E}[Y^2]$. En conclure que X et Y ne sont pas indépendantes. Le vecteur (X,Y) est-il gaussien ? △

Proposition VI.7. *On considère une transformation affine de $\mathbb{R}^d$ dans $\mathbb{R}^n$: $x\mapsto Mx+T$, où M est une matrice déterministe de taille $n\times d$ et T un vecteur déterministe de $\mathbb{R}^n$. Soit X un vecteur gaussien à valeurs dans $\mathbb{R}^d$ et de loi $\mathcal{N}(\mu,V)$. La variable aléatoire $MX+T$ est un vecteur gaussien à valeurs dans $\mathbb{R}^n$. De plus sa loi est :*

$$\boxed{\mathcal{L}(MX+T)=\mathcal{N}(T+M\mu, MVM^t).}$$

Démonstration. On pose $Y = MX + T$. Il s'agit d'un vecteur aléatoire à valeurs dans $\mathbb{R}^n$. Soit $u \in \mathbb{R}^n$. On a :

$$\psi_Y(u) = \mathbb{E}\left[\mathrm{e}^{i\langle u,Y\rangle}\right] = \mathbb{E}\left[\mathrm{e}^{i\langle u,MX+T\rangle}\right] = \mathrm{e}^{i\langle u,T\rangle}\,\mathbb{E}\left[\mathrm{e}^{i\langle M^t u,X\rangle}\right],$$

où M^t est la transposée de la matrice M. Grâce à la proposition VI.5, il vient :

$$\begin{aligned}\psi_Y(u) &= \exp\left(i\langle u,T\rangle + i\langle M^t u,\mu\rangle - \frac{\langle M^t u, VM^t u\rangle}{2}\right)\\ &= \exp\left(i\langle u,T+M\mu\rangle - \frac{\langle u, MVM^t u\rangle}{2}\right).\end{aligned}$$

Ceci étant vrai pour tout $u \in \mathbb{R}^n$, on en déduit que $\mathcal{L}(Y) = \mathcal{N}(T + M\mu, MVM^t)$. □

Exercice VI.2.
Soit X un vecteur gaussien de loi $\mathcal{N}(\mu, I_d)$, où I_d est la matrice identité de taille $d\times d$. Soit P_1 et P_2 deux projections orthogonales sur des sous-espaces orthogonaux de $\mathbb{R}^d$ ($P_1P_2 = P_2P_1 = 0$). Montrer que P_1X et P_2X sont indépendants. Donner la loi du couple (P_1X, P_2X). △

Théorème VI.8. *Soit V une matrice symétrique positive de taille $d \times d$, et μ un vecteur de $\mathbb{R}^d$. Il existe un vecteur gaussien de loi $\mathcal{N}(\mu, V)$.*

Démonstration. Soit Y un vecteur aléatoire de $\mathbb{R}^d$ composé de v.a. réelles indépendantes de loi $\mathcal{N}(0,1)$. L'exemple VI.3 assure que $\mathcal{L}(Y) = \mathcal{N}(0, I_d)$, où I_d est la matrice identité de taille $d\times d$. Comme V est symétrique réelle, il existe une matrice diagonale $D = \mathrm{Diag}\,(d_1, \ldots, d_d)$ et une matrice orthogonale U ($U^tU = UU^t = I_d$) telles que $V = UDU^t$. La matrice V étant positive, cela entraîne que $d_k \geq 0$. On pose $\sigma_k = \sqrt{d_k}$ et la matrice diagonale $\Delta = \mathrm{Diag}\,(\sigma_1, \ldots, \sigma_d)$. Donc on a $V = U\Delta\Delta U^t$. Le vecteur aléatoire $X = \mu + U\Delta Y$ est un vecteur gaussien de loi $\mathcal{N}(\mu + U\Delta 0, U\Delta\Delta^t U^t) = \mathcal{N}(\mu, V)$. □

Si $\det V = 0$, alors il est facile de déduire de la démonstration précédente qu'il existe une combinaison linéaire non triviale du vecteur X qui est p.s. constante. On parle de vecteur gaussien dégénéré.

Proposition VI.9. *Soit X un vecteur gaussien de $\mathbb{R}^d$ non dégénéré ($\det V > 0$) de loi $\mathcal{N}(\mu, V)$. Alors la matrice V est inversible et la loi de X possède la densité f : pour $x \in \mathbb{R}^d$,*

$$\boxed{f_X(x) = \frac{1}{(2\pi)^{d/2}(\det V)^{1/2}} \exp\left(-\frac{\langle x-\mu, V^{-1}(x-\mu)\rangle}{2}\right).}$$

Démonstration. On reprend les notations de la démonstration du théorème précédent. Par indépendance, la densité de la loi du vecteur Y s'écrit comme le produit des densités de chacune des composantes :

$$f_Y(y_1, \dots, y_d) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{{y_k}^2}{2}} = \frac{1}{(\sqrt{2\pi})^{d/2}} \mathrm{e}^{-\frac{|y|^2}{2}}.$$

On utilise la méthode de la fonction muette. Soit g mesurable bornée définie sur $\mathbb{R}^d$. On a :

$$\mathbb{E}[g(X)] = \mathbb{E}[g(\mu + U\Delta Y)] = \int_{\mathbb{R}^d} g(\mu + U\Delta y)\, \mathrm{e}^{-\frac{|y|^2}{2}} \frac{dy}{(2\pi)^{d/2}}.$$

Les deux conditions $V = U\Delta^2 U^t$ et $\det V > 0$ impliquent que la matrice diagonale $\Delta = \mathrm{Diag}\ (\sigma_1, \dots, \sigma_d)$ est inversible. En particulier $\sigma_i > 0$ pour $i \in \{1, \dots, d\}$. On effectue le changement de variable suivant : $x = \varphi(y) = \mu + U\Delta y$, soit $y = \Delta^{-1}U^t(x-\mu)$. Ainsi on a :

$$\langle y, y\rangle = \langle \Delta^{-1}U^t(x-\mu), \Delta^{-1}U^t(x-\mu)\rangle = \langle x-\mu, V^{-1}(x-\mu)\rangle.$$

Comme Δ est inversible, φ est un difféomorphisme de classe C^1 de $\mathbb{R}^d$ dans $\mathbb{R}^d$. Et on a :

$$|\mathrm{Jac}[\varphi](x)| = |\det U\Delta| = |\det \Delta|\,.$$

On remarque que $\det V = \det U\Delta^2 U^t = (\det \Delta)^2$ et donc $|\mathrm{Jac}[\varphi](x)| = \sqrt{\det V}$. Il vient :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x) \frac{1}{(2\pi)^{d/2}(\det V)^{1/2}} \mathrm{e}^{-\frac{\langle x-\mu, V^{-1}(x-\mu)\rangle}{2}}\, dx.$$

On en déduit la densité de la loi de X. □

Application. Soit (X,Y) un vecteur gaussien de $\mathbb{R}^2$. On suppose que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, et $\mathrm{Var}(X) = \sigma_1^2$, $\mathrm{Var}(Y) = \sigma_2^2$, $\mathrm{Cov}(X,Y) = \rho$. La matrice de covariance du vecteur (X,Y) est $V = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$. On a $\det V = \sigma_1^2\sigma_2^2 - \rho^2$ et si $\rho^2 < \sigma_1^2\sigma_2^2$, $V^{-1} = \dfrac{1}{\sigma_1^2\sigma_2^2 - \rho^2} \begin{pmatrix} \sigma_2^2 & -\rho \\ -\rho & \sigma_1^2 \end{pmatrix}$.

Le vecteur est dégénéré si $\det V = 0$, c'est-à-dire si $\sigma_1^2\sigma_2^2 = \rho^2$. Dans ce cas, on suppose par exemple que ρ est positif. On remarque alors que :

$$\begin{aligned} \mathrm{Var}(\sigma_2 X - \sigma_1 Y) &= \sigma_2^2 \,\mathrm{Var}(X) + \sigma_1^2 \,\mathrm{Var}(Y) - 2\sigma_1\sigma_2 \,\mathrm{Cov}(X,Y) \\ &= 2\sigma_1\sigma_2(\sigma_1\sigma_2 - \rho) \\ &= 0. \end{aligned}$$

La v.a. gaussienne $\sigma_2 X - \sigma_1 Y$ est donc constante : elle est dégénérée.

Dans le cas non dégénéré, $\det V > 0$, on obtient la densité f du vecteur (X,Y) à l'aide de la proposition précédente :

$$f(x,y) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \rho^2}} \exp\left(-\frac{\sigma_2^2 x^2 + \sigma_1^2 y^2 - 2\rho xy}{2(\sigma_1^2\sigma_2^2 - \rho^2)} \right).$$

Si $\rho = 0$, on vérifie bien que $f(x,y) = f_{\sigma_1^2}(x) f_{\sigma_2^2}(y)$, où f_{σ^2} est la densité de la loi gaussienne réelle $\mathcal{N}(0,\sigma^2)$.

Dans la figure VI.1, on présente plusieurs réalisations d'un couple gaussien en fonction de ρ. ◊

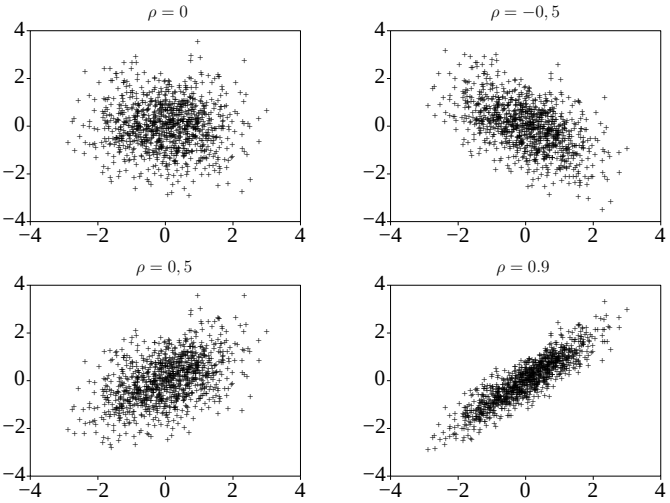


Figure VI.1. 1000 réalisations d'un vecteur gaussien (X,Y), où $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ et $\mathrm{Cov}(X,Y) = \rho$.

Une propriété importante des vecteurs gaussiens concerne les lois conditionnelles. Si (X, Y) est un vecteur gaussien, alors la loi de X sachant Y est une loi gaussienne dont les paramètres dépendent de Y, et des matrices de covariance.

Proposition VI.10. *Soit (X, Y) un vecteur gaussien (de dimension $d + n$). On suppose que la matrice de covariance $\mathrm{Cov}(Y, Y)$ est inversible (le vecteur Y est non dégénéré). La loi conditionnelle de X sachant Y est une loi gaussienne $\mathcal{N}(\mathbb{E}[X|Y], \Sigma)$. De plus, on a :*

$$\mathbb{E}[X|Y] = \mathbb{E}[X] + \mathrm{Cov}(X, Y)\,\mathrm{Cov}(Y, Y)^{-1}(Y - \mathbb{E}[Y])$$
$$\textit{et} \quad \Sigma = \mathrm{Cov}(X, X) - \mathrm{Cov}(X, Y)\,\mathrm{Cov}(Y, Y)^{-1}\,\mathrm{Cov}(Y, X).$$

On remarque que la matrice de covariance Σ est déterministe, alors que l'espérance conditionnelle $\mathbb{E}[X|Y]$ est aléatoire. Cette dernière est une fonction linéaire de Y.

Démonstration. On pose $M = \mathrm{Cov}(X, Y)\,\mathrm{Cov}(Y, Y)^{-1}$ et $W = X - MY$. Le vecteur (W, Y) est gaussien (il s'agit d'une transformation linéaire du vecteur gaussien (X, Y)). Le vecteur W est un vecteur gaussien indépendant de Y car $\mathrm{Cov}(W, Y) = \mathrm{Cov}(X, X) - M\,\mathrm{Cov}(Y, Y) = 0$. On détermine les paramètres de sa loi :

$$\begin{aligned}\mathbb{E}[W] &= \mathbb{E}[X - MY] = \mathbb{E}[X] - \mathrm{Cov}(X, X)\,\mathrm{Cov}(Y, Y)^{-1}\mathbb{E}[Y],\\ \mathrm{Cov}(W, W) &= \mathrm{Cov}(X, X) - M\,\mathrm{Cov}(Y, X)\\ &= \mathrm{Cov}(X, X) - \mathrm{Cov}(X, Y)\,\mathrm{Cov}(Y, Y)^{-1}\,\mathrm{Cov}(Y, X).\end{aligned}$$

Comme W est indépendant de Y, la loi conditionnelle de $X = MY + W$ sachant Y est celle de $a + W$, avec $a = MY$. La loi conditionnelle de X sachant Y, $\mathcal{L}(X|Y)$, est donc une loi gaussienne d'espérance $\mathbb{E}[W] + MY$ et de matrice de covariance $\Sigma = \mathrm{Cov}(W, W)$. □

VI.2 Loi du χ^2, loi de Student, loi de Fisher

Un certain nombre de lois sont naturellement associées aux lois gaussiennes. On rappelle que les lois du χ^2 ont été définies au chapitre III.5, comme cas particuliers des lois Gamma.

Exercice VI.3.
Soit une suite $(X_k, k \in \mathbb{N}^*)$ de variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. Montrer que la loi de X_k^2 est la loi du χ^2 à 1 degré de liberté. Montrer que la loi de $\sum_{k=1}^n X_k^2$ est la loi du χ^2 à n degrés de liberté.

△

Correction VI.3. On utilise la méthode de la fonction muette pour déterminer la loi de X_n^2. Soit g une fonction mesurable bornée. On a :

$$\mathbb{E}[g(X_k^2)] = \int_{\mathbb{R}} g(x^2)\,\mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}} = 2\int_0^\infty g(x^2)\,\mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}.$$

On fait le changement de variable $y = x^2$ sur $]0,+\infty[$. Il vient :

$$\mathbb{E}[g(X_k^2)] = \int_0^\infty g(y)\,\frac{1}{\sqrt{2\pi}}\,y^{-1/2}\,\mathrm{e}^{-y/2}\ dy.$$

On en déduit que la loi de X_k^2 est la loi Gamma $\Gamma(1/2, 1/2)$. Il s'agit bien de la loi $\chi^2(1)$.

Les variables aléatoires $(X_k, k \in \mathbb{N}^*)$ étant indépendantes, on en déduit que les variables aléatoires $(X_k^2, k \in \mathbb{N}^*)$ sont indépendantes. La fonction caractéristique de $S_n = \sum_{k=1}^n X_k^2$ est, par indépendance, $\psi_{S_n}(u) = \left(\dfrac{1}{1-2iu}\right)^{n/2}$. On reconnaît la fonction caractéristique de la loi $\chi^2(n)$. ▲

Le théorème de Cochran qui suit est utile en statistique.

Théorème VI.11. *Soit $X = (X_1, \ldots, X_n)$ un vecteur de variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. Soit $E_1, \ldots, E_p$ une famille de $p \geq 2$ sous-espaces vectoriels de $\mathbb{R}^n$ orthogonaux deux à deux tels que $E_1 \oplus \cdots \oplus E_p = \mathbb{R}^n$. On suppose que la dimension de E_i, n_i, est non nulle. On note X_{E_i} la projection orthogonale de X sur E_i. Alors, les variables $X_{E_1}, \ldots, X_{E_p}$ sont indépendantes et la loi de $|X_{E_i}|^2$ est une loi du χ^2 de paramètre n_i.*

Démonstration. Soit $e = \{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$. Soit une base orthonormée de $\mathbb{R}^n$, $f = \{f_1, \ldots, f_n\}$, et $Y = (Y_1, \ldots, Y_n)$ les coordonnées de X dans la base f. Il existe une matrice U de taille $n \times n$ telle que si $x = (x_1, \ldots, x_n)$ sont les coordonnées d'un vecteur dans la base e, alors ses coordonnées dans la base f sont données par $y = Ux$. De plus on a $U^tU = UU^t = I_n$, où I_n est la matrice

identité. On a donc $Y = UX$. Comme X est un vecteur gaussien de moyenne nulle et de matrice de covariance la matrice identité, I_n, on en déduit que Y est un vecteur gaussien de moyenne $U\mathbb{E}[X] = 0$ et de matrice de covariance $UI_nU^t = I_n$. Ainsi X et Y ont même loi.

On note $f^{(i)} = \{f_1^{(i)}, \dots, f_{n_i}^{(i)}\}$ une base orthonormée de E_i. Ainsi $f = \cup_{1\leq i\leq p} f^{(i)}$ est une base orthonormée de $\mathbb{R}^n$. Soit $Y = (Y_1, \dots, Y_n)$ les coordonnées du vecteur X dans la base f. Ainsi on a $X_{E_i} = \sum_{j=1}^{n_i} Y_{m_i+j} f_{m_i+j}$, où $m_i = 0$ si $i = 1$ et $m_i = \sum_{k=1}^{i-1} n_k$ sinon. D'après ce qui précède, les variables $Y_1, \dots, Y_n$ sont indépendantes de loi $\mathcal{N}(0, 1)$. On en déduit donc que les variables $X_{E_1}, \dots, X_{E_p}$ sont indépendantes. On a également :

$$|X_{E_i}|^2 = \sum_{j=1}^{n_i} Y_{m_i+j}^2.$$

On en déduit que $|X_{E_i}|^2$ est la somme de n_i carrés de gaussiennes centrées réduites indépendantes. Sa loi est donc la loi du χ^2 à n_i degrés de liberté. □

On donne une application importante du théorème de Cochran. Soit $\mu \in \mathbb{R}$ et $\sigma > 0$. On considère une suite $(X_k, k \in \mathbb{N}^*)$ de variables aléatoires indépendantes et de loi $\mathcal{N}(\mu, \sigma^2)$. On pose :

$$\boxed{\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k,} \quad \text{et} \quad \boxed{V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2.}$$

On peut réécrire V_n comme la variance empirique, à un coefficient multiplicatif près :

$$\boxed{V_n = \frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}_n^2\right).}$$

Par la loi forte des grands nombres, on en déduit que le couple $(\bar{X}_n, V_n)$ converge p.s. vers (μ, σ^2). On donne en fait la loi du couple $(\bar{X}_n, V_n)$.

Proposition VI.12. *Les variables $\bar{X}_n$ et V_n sont indépendantes. La loi de $\bar{X}_n$ est la loi gaussienne $\mathcal{N}(\mu, \sigma^2/n)$. La loi de $(n-1)V_n/\sigma^2$ est la loi du χ^2 à $n-1$ degrés de liberté.*

Démonstration. Soit $X = (X_1, \dots, X_n)$. On note $\mathbf{1}_n \in \mathbb{R}^n$ le vecteur dont toutes les coordonnées sont égales à 1. On pose $X = \mu\mathbf{1}_n + \sigma\varepsilon$. Le vecteur $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$

est composé de variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. On note Δ la droite vectorielle engendrée par le vecteur $\mathbf{1}_n$. On note H le sous-espace vectoriel orthogonal à Δ (en particulier $\Delta \oplus H = \mathbb{R}^n$). La projection orthogonale de X sur Δ est $X_\Delta = \mu \mathbf{1}_n + \sigma \varepsilon_\Delta = \bar{X}_n \mathbf{1}_n$, où ε_Δ est la projection orthogonale de ε sur Δ. On note ε_H la projection orthogonale de ε sur H. On a :

$$\varepsilon_H = \varepsilon - \varepsilon_\Delta = \frac{X - X_\Delta}{\sigma} = \frac{X - \bar{X}_n \mathbf{1}_n}{\sigma}.$$

On remarque que $(n-1)V_n = |X - \bar{X}_n \mathbf{1}_n|^2 = \sigma^2 \, |\varepsilon_H|^2$.

Le théorème de Cochran assure que les vecteurs ε_Δ et ε_H sont indépendants et que la loi de $|\varepsilon_H|^2$ est la loi du χ^2 de paramètre $n-1$. On en déduit donc que les vecteurs $\dfrac{\bar{X}_n - \mu}{\sigma}\mathbf{1}_n$ et $\dfrac{X - \bar{X}_n \mathbf{1}_n}{\sigma}$, et donc les variables $\bar{X}_n$ et V_n, sont indépendants. De plus, $(n-1)V_n/\sigma^2$ suit la loi d'un χ^2 de paramètre $n-1$. Enfin, comme $\bar{X}_n$ est une combinaison linéaire des composantes du vecteur gaussien X, c'est donc une variable aléatoire gaussienne. On détermine les paramètres de sa loi. On a $\mathbb{E}[\bar{X}_n] = \mu$ par linéarité, et $\mathrm{Var}(\bar{X}_n) = \dfrac{\sigma^2}{n}$ en utilisant l'indépendance. La loi de $\bar{X}_n$ est donc $\mathcal{N}(\mu, \sigma^2/n)$. □

On déduit des propriétés de la loi du χ^2 que :

$$\mathbb{E}[V_n] = \sigma^2 \quad \text{et} \quad \mathrm{Var}(V_n) = 2\sigma^4/(n-1).$$

On déduit de la loi forte des grands nombres que $\bar{X}_n$ converge p.s. vers μ et V_n converge p.s. vers σ^2. On dit que $\bar{X}_n$ et V_n sont des **estimateurs convergents** de μ et σ^2. Comme $\mathbb{E}[\bar{X}_n] = \mu$ et $\mathbb{E}[V_n] = \sigma^2$, on parle d'estimateurs **sans biais**. C'est la raison de la normalisation par $(n-1)$ et non par n dans la définition de V_n. Les estimateurs sont utilisés pour donner une estimation et un intervalle de confiance sur μ et σ^2 quand ceux-ci sont inconnus, voir les exercices X.1 et X.3.

Définition VI.13. *Si G et U sont deux v.a. indépendantes et de lois respectives $\mathcal{N}(0,1)$ et $\chi^2(n)$, alors on dit que la variable aléatoire :*

$$\boxed{T_n = \frac{G}{\left[\dfrac{U}{n}\right]^{1/2}}}$$

*suit une loi de **Student** de paramètre n.*

La fonction de répartition de la loi de Student est tabulée (voir le paragraphe XI.4). Voir la figure VI.2 pour la densité de la loi de Student ainsi que l'exercice ci-dessous.

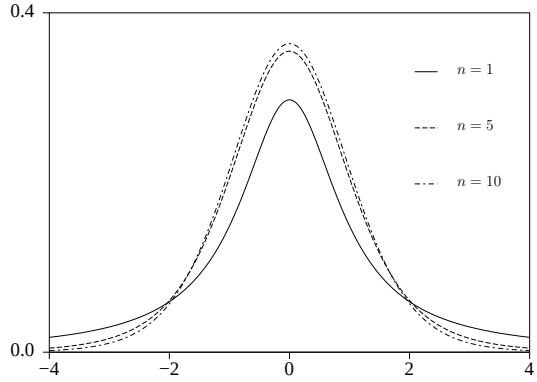


Figure VI.2. Densités de la loi de Student de paramètre n.

Exercice VI.4.
Montrer que la densité de la loi de Student de paramètre n est :

$$f_{T_n}(t) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \quad t \in \mathbb{R}.$$

△

Exercice VI.5.
Montrer que la suite $(T_n, n \in \mathbb{N}^*)$, où T_n est de loi de Student de paramètre n, converge en loi vers la loi gaussienne $\mathcal{N}(0,1)$. △

Correction VI.5. Soit $(G_n, n \in \mathbb{N})$ une suite de v.a. indépendantes de loi gaussienne $\mathcal{N}(0,1)$. Par définition T_n a même loi que $T'_n = G_0/\sigma_n$ où $\sigma_n^2 = \frac{1}{n}\sum_{k=1}^n G_k^2$. La loi forte des grands nombres assure que $(\sigma_n^2, n \in \mathbb{N}^*)$ converge p.s. vers 1. On en déduit que la suite $(T'_n, n \in \mathbb{N}^*)$ converge p.s. vers G_0. Comme la suite $(T_n, n \in \mathbb{N}^*)$ a même loi que $(T'_n, n \in \mathbb{N}^*)$, on en déduit qu'elle converge en loi vers G_0. ▲

Afin d'avoir des quantités normalisées, on pose :

$$G = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \quad \text{et} \quad U = \frac{n-1}{\sigma^2} V_n.$$

On a vu que les v.a. U et V sont indépendantes et de lois respectives $\chi^2(n-1)$ et $\mathcal{N}(0,1)$. La variable aléatoire :

$$\boxed{T_{n-1} = \frac{G}{\left[\frac{U}{(n-1)}\right]^{1/2}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{V_n}},}$$

suit donc une loi de Student de paramètre $n-1$. En particulier la loi de T_{n-1} ne dépend pas des paramètres μ et σ^2.

Exemple VI.14. Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes de même loi gaussienne dont ne connaît pas les paramètres. La moyenne empirique fournit une estimation de la moyenne μ. De plus on connaît exactement le niveau de l'intervalle de confiance :

$$I_n = \left[\bar{X}_n - \frac{a\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \frac{a\sqrt{V_n}}{\sqrt{n}}\right].$$

En effet, on a $\mathbb{P}(\mu \in I_n) = \mathbb{P}(|T_{n-1}| \leq a)$. ◊

Enfin on présente une famille de lois qui joue un rôle important dans les modèles linéaires gaussiens.

Définition VI.15. *Si U et V sont deux v.a. indépendantes de loi $\chi^2(n)$ et $\chi^2(m)$, alors on dit que* $\boxed{S = \frac{U}{n}\frac{m}{V}}$ *suit une loi de* ***Fisher-Snedecor*** *de paramètre (n,m).*

La fonction de répartition de la loi de Fisher-Snedecor est tabulée (voir le paragraphe XI.5). Voir la figure VI.3 pour la densité de la loi de Fisher-Snedecor ainsi que l'exercice ci-dessous.

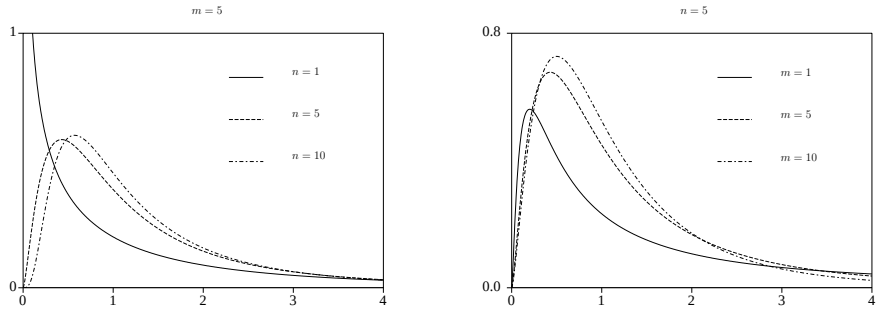


Figure VI.3. Densités de la loi de Fisher-Snedecor de paramètre (n,m).

Exercice VI.6.
Montrer que la densité de la loi de Fisher-Snedecor de paramètre (n, m) est :

$$f_S(s) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} \frac{s^{\frac{n}{2}-1}}{\left(1+\frac{sn}{m}\right)^{(n+m)/2}} \mathbf{1}_{\{s\geq 0\}}.$$

△

VI.3 Théorème central limite vectoriel

On généralise le théorème central limite au cas multidimensionnel.

Proposition VI.16. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$, indépendantes de même loi et de carré intégrable. On pose $\mu = \mathbb{E}[X_1] \in \mathbb{R}^d$, $V = \mathrm{Cov}(X_1, X_1) \in \mathbb{R}^{d\times d}$ et $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ la moyenne empirique. Alors, la suite de vecteurs aléatoires $(\sqrt{n}(\bar{X}_n - \mu), n \in \mathbb{N}^*)$ converge en loi vers le vecteur gaussien de loi $\mathcal{N}(0, V)$:*

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} \mathcal{N}(0, V).}$$

Démonstration. Soit $u \in \mathbb{R}^d$. On a $\psi_{\sqrt{n}(\bar{X}_n-\mu)}(u) = \psi_{\sqrt{n}\langle \bar{X}_n-\mu, u\rangle}(1)$. On pose $Z_k = \langle X_k - \mu, u\rangle$. Les v.a. $(Z_k, k \in \mathbb{N}^*)$ sont indépendantes, de même loi et de carré intégrable. On a $\mathbb{E}[Z_k] = 0$ et $\mathrm{Var}(Z_k) = \langle u, Vu\rangle$. On déduit du théorème central limite en dimension 1, que la suite $(\sqrt{n}\,\frac{1}{n}\sum_{k=1}^n Z_k, n \in \mathbb{N}^*)$ converge en loi vers $\mathcal{N}(0, \langle u, Vu\rangle)$. On en déduit que :

$$\lim_{n\to\infty} \psi_{\sqrt{n}(\bar{X}_n-\mu)}(u) = \lim_{n\to\infty} \psi_{\sqrt{n}\,\frac{1}{n}\sum_{k=1}^n Z_k}(1) = \mathrm{e}^{-\langle u, Vu\rangle/2}.$$

Comme ceci est vrai pour tout $u \in \mathbb{R}^d$, on en déduit la proposition. □

Le corollaire suivant est très important en pratique pour construire les intervalles de confiance.

Corollaire VI.17. *Soit $(X_k, k \in \mathbb{N}^*)$ une suite de v.a. à valeurs dans $\mathbb{R}^d$ indépendantes de même loi et de carré intégrable. On pose $\mu = \mathbb{E}[X_1] \in \mathbb{R}^d$, $V = \mathrm{Cov}(X_1, X_1) \in \mathbb{R}^{d\times d}$ et $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ la moyenne empirique. Soit g une*

fonction mesurable de $\mathbb{R}^d$ dans $\mathbb{R}^p$ continue et différentiable en μ. Sa différentielle au point μ est la matrice $\dfrac{\partial g}{\partial x}(\mu)$ de taille $p \times d$ définie par :

$$\left(\frac{\partial g}{\partial x}(\mu)\right)_{i,j} = \frac{\partial g_i}{\partial x_j}(\mu); \quad 1 \le i \le p, \quad 1 \le j \le d.$$

On pose $\Sigma = \dfrac{\partial g}{\partial x}(\mu)\, V\, \left(\dfrac{\partial g}{\partial x}(\mu)\right)^t$. On a alors :

$$\boxed{g(\bar{X}_n) \xrightarrow[n\to\infty]{\textbf{\textit{p.s.}}} g(\mu) \quad \textit{et} \quad \sqrt{n}\Big(g(\bar{X}_n) - g(\mu)\Big) \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} \mathcal{N}(0, \Sigma).}$$

Démonstration. La convergence p.s. de la suite $(g(\bar{X}_n), n \in \mathbb{N}^*)$ est une conséquence de la loi forte des grands nombres et de la continuité de g en μ. Pour $x = (x_1, \ldots, x_d)$, $i \in \{0, \ldots, d\}$, on note $x^{(i)} = (y_1, \ldots, y_d)$ le vecteur tel que $y_j = x_j$ si $j \le i$ et $y_j = \mu_j$ si $j > i$. En particulier $x^{(0)} = \mu$ et $x^{(d)} = x$. La matrice $h(x)$ de taille $p \times d$ définie par :

$$h(x)_{i,j} = \begin{cases} \dfrac{g_i(x^{(i)}) - g_i(x^{(i-1)})}{x_j - \mu_j} & \text{si } x_j \neq \mu_j, \\ \dfrac{\partial g_i}{\partial x_j}(x^{(i-1)}) & \text{si } x_j = \mu_j \end{cases}$$

est une fonction continue de x en μ. On déduit du lemme V.2 que $(h(\bar{X}_n), n \in \mathbb{N}^*)$ converge p.s. vers $h(\mu) = \dfrac{\partial g}{\partial x}(\mu)$. On déduit du théorème de Slutsky que la suite $\Big((h(\bar{X}_n), \sqrt{n}(\bar{X}_n - \mu)), n \in \mathbb{N}^*\Big)$ converge en loi vers $\left(\dfrac{\partial g}{\partial x}(\mu), Z\right)$, où $\mathcal{L}(Z) = \mathcal{N}(0, V)$. Enfin la fonction $(x, y) \mapsto xy$ est continue. On déduit du corollaire V.21, que $\Big(h(\bar{X}_n)\sqrt{n}(\bar{X}_n - \mu), n \in \mathbb{N}^*\Big)$ converge en loi vers $\dfrac{\partial g}{\partial x}(\mu)Z$. Pour conclure, on remarque que $h(\bar{X}_n)\sqrt{n}(\bar{X}_n - \mu) = \sqrt{n}(g(\bar{X}_n) - g(\mu))$ et la loi de $\dfrac{\partial g}{\partial x}(\mu)Z$ est d'après la proposition VI.7 la loi gaussienne $\mathcal{N}(0, \Sigma)$. □

Application. On considère une suite $X_1, \ldots, X_n$ de v.a. indépendantes de loi de Poisson de paramètre θ inconnu. On rappelle que $\mathbb{E}[X_1] = \theta$ et $\mathrm{Var}(X_1) = \theta$. On dispose d'une réalisation de cette suite, et on désire estimer θ. La loi forte des grands nombres assure que la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ converge p.s. vers $\mathbb{E}[X_1] = \theta$. De plus par le théorème central limite, on sait que $\sqrt{n}(\bar{X}_n - \theta)$

converge en loi vers $\mathcal{N}(0,\theta)$. La loi limite dépend de θ, on ne peut donc pas donner directement d'intervalle de confiance pour θ. On peut diviser $\sqrt{n}(\bar{X}_n - \theta)$ par la variance empirique, où tout autre v.a. qui converge p.s. (ou même en loi) vers $\sqrt{\theta}$. La méthode de **stabilisation de la variance** consiste à chercher une fonction g régulière telle que $g'(\mathbb{E}[X_1])^2 \operatorname{Var}(X_1) = 1$. On déduit alors du corollaire précédent que $\sqrt{n}(g(\bar{X}_n) - g(\theta))$ converge en loi vers $\mathcal{N}(0,1)$. La limite est alors indépendante du paramètre inconnu. Pour les v.a. de Poisson, la fonction g doit être solution de $g'(\theta)^2\theta = 1$ soit $g(\theta) = 2\sqrt{\theta}$. Ainsi $2\sqrt{\theta}$ appartient à l'intervalle $\left[2\sqrt{\bar{X}_n} - 1,96/\sqrt{n}; 2\sqrt{\bar{X}_n} + 1,96/\sqrt{n}\right]$ avec une probabilité asymptotique de 95%. Autrement dit $\left[\left(\sqrt{\bar{X}_n} - \frac{1,96}{2\sqrt{n}}\right)^2, \left(\sqrt{\bar{X}_n} + \frac{1,96}{2\sqrt{n}}\right)^2\right]$ est un intervalle de confiance pour θ de niveau asymptotique 95%. ◊

VI.4 Résumé

- Une v.a. vectorielle est un **vecteur gaussien** si et seulement si toute combinaison linéaire de ses coordonnées est une v.a. gaussienne, éventuellement dégénérée (i.e. de variance nulle).
- Toute limite en loi de vecteurs gaussiens est un vecteur gaussien.
- Toute transformation affine d'un vecteur gaussien est un vecteur gaussien.
- La loi d'un vecteur gaussien est caractérisée par sa **moyenne** et sa **matrice de covariance**.
- Le vecteur gaussien est non dégénéré si et seulement si le déterminant de sa matrice de covariance est strictement positif.
- Si (X,Y) est un vecteur gaussien, alors X et Y sont **indépendants** si et seulement si la matrice $\operatorname{Cov}(X,Y)$ est nulle. L'hypothèse "**vecteur gaussien**" est cruciale.
- Soit $(X_k, k \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes de loi $\mathcal{N}(\mu,\sigma^2)$. La **moyenne empirique** $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ et la **variance empirique** $V_n = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ sont des **estimateurs convergents** et **sans biais** de μ et σ^2. De plus $\bar{X}_n$ et V_n sont indépendants. On a $\mathcal{L}(\bar{X}_n) = \mathcal{N}(\mu, \sigma^2/n)$ et $\mathcal{L}((n-1)V_n/\sigma^2) = \chi^2(n-1)$.

– Soit $(X_k, k \in \mathbb{N}^*)$ une suite de v.a. vectorielles indépendantes et de même loi. On suppose que X_k est de carré intégrable, et on note μ sa moyenne et V sa matrice de covariance. Alors on a :

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\textbf{en loi}} \mathcal{N}(0, V).}$$

– Soit g une application continue et différentiable en μ.
On pose $\Sigma = \dfrac{\partial g}{\partial x}(\mu)\ V\ \left(\dfrac{\partial g}{\partial x}(\mu)\right)^t$. On a alors :

$$\boxed{g(\bar{X}_n) \xrightarrow[n\to\infty]{\textbf{p.s.}} g(\mu) \quad \text{et} \quad \sqrt{n}\Big(g(\bar{X}_n) - g(\mu)\Big) \xrightarrow[n\to\infty]{\textbf{en loi}} \mathcal{N}(0, \Sigma).}$$

VI.5 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice VI.1 p. 169, *Exercice* VI.2 p. 170, *Exercice* VI.3 p. 174, *Exercice* VI.4 p. 177, *Exercice* VI.5 p. 177, *Exercice* VI.6 p. 179.

Exercice VI.7.
Soit X, Y deux variables aléatoires réelles indépendantes de loi $\mathcal{N}(0,1)$. Calculer la loi conditionnelle de X sachant $X+Y$. Calculer la loi conditionnelle de $X+Y$ sachant $X-Y$. △

Exercice VI.8.
Soit (X,Y) un vecteur gaussien tel que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ et $\mathrm{Cov}(X,Y) = \cos\phi$. On définit $R = \sqrt{X^2+Y^2}$ et Θ par $R\cos\Theta = X$, $R\sin\Theta = Y$.

1. Donner la densité de la loi de (X,Y) quand elle existe.
2. Calculer la loi de (R,Θ). Calculer la loi de Θ.
3. Calculer $\mathbb{P}(XY > 0)$.

△

Exercice VI.9.
Soit X_1, X_2 deux variables aléatoires indépendantes de loi $\mathcal{N}(0,\sigma^2)$ (avec $\sigma > 0$).

1. Donner en utilisant les coordonnées polaires, la loi de (Y,Z), où :
$$Y = \frac{X_1^2 - X_2^2}{\sqrt{X_1^2+X_2^2}} \quad \text{et} \quad Z = \frac{2X_1X_2}{\sqrt{X_1^2+X_2^2}}.$$
2. Donner la loi jointe du signe de X_1X_2 et du signe de $X_1^2 - X_2^2$.

△

Exercice VI.10.
Soit $(V_n, n \geq 1)$ une suite de vecteurs gaussiens de dimension d et de loi $\mathcal{N}(B_n, A_n)$ telle que la suite de vecteurs $(B_n, n \in \mathbb{N}^*)$ et la suite de matrices $(A_n, n \in \mathbb{N}^*)$ convergent. Montrer que $(V_n, n \geq 1)$ converge en loi. Donner la loi limite. △

Exercice VI.11.
Soit $(Y_n, n \geq 1)$ des variables aléatoires indépendantes et de loi $\mathcal{N}(\mu,\sigma^2)$. Soit $a \in \mathbb{R}$. Soit X_0 une v.a. réelle de loi $\mathcal{N}(\mu_0,\sigma_0^2)$, indépendante de $(Y_n, n \geq 1)$. Pour

tout $n \geq 1$, on pose $X_n = aX_{n-1} + Y_n$. On modélise ainsi des résultats qui à l'étape n dépendent linéairement des résultats à l'étape $n-1$ mais qui sont bruités.

1. Montrer que $(X_0, \cdots, X_n)$ est un vecteur gaussien. Déterminer la loi de X_n et calculer $\mathrm{Cov}(X_k, X_n)$ pour $0 \leq k \leq n$.
2. À quelle condition sur a, la suite $(X_n, n \in \mathbb{N})$ converge-t-elle en loi ? Déterminer la loi limite. Quelle est la loi de X_1, puis celle de X_n si la loi de X_0 est la loi limite ?
3. Montrer que si $a \in]-1, 1[$, alors le vecteur (X_n, X_{n+1}) converge en loi vers un vecteur gaussien dont on déterminera les paramètres.
4. Étudier la convergence en loi de la moyenne empirique $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

△

Exercice VI.12.
Soit (X, Y) un vecteur aléatoire de loi uniforme sur le disque unité de $\mathbb{R}^2$: sa loi possède la densité $f(x, y) = \dfrac{1}{\pi}\,\mathbf{1}_{\{x^2+y^2\leq 1\}}$.

1. Calculer la loi de X.
2. Vérifier que X et Y ne sont pas indépendants, mais que $\mathrm{Cov}(X, Y) = 0$.

Soit $((X_n, Y_n), n \in \mathbb{N}^*)$ une suite de v.a. indépendantes de même loi que (X, Y). On pose $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ et $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$.

3. Montrer en utilisant les fonctions caractéristiques que $\bar{X}_n$ et $\bar{Y}_n$ ne sont pas indépendants.
4. Montrer que le couple $(\sqrt{n}\bar{X}_n, \sqrt{n}\bar{Y}_n)$ converge en loi vers un couple de v.a. indépendantes.

△

Exercice VI.13.
Soit $(X_k, k \in \mathbb{N}^*)$ une suite de variables aléatoires telles que X_k est de loi $\chi^2(k)$. Montrer que la suite $\left(\sqrt{2X_k} - \sqrt{2k-1}, k \in \mathbb{N}^*\right)$ converge en loi vers une loi gaussienne. On pourra utiliser le fait que X_k a même loi que $\sum_{i=1}^{k} Y_i^2$, où les variables aléatoires Y_i sont indépendantes de loi $\mathcal{N}(0, 1)$. △

Exercice VI.14.
Soit $X_1, \ldots, X_n$ des variables aléatoires réelles indépendantes de même loi, de carré intégrable, d'espérance m, de variance σ^2. On suppose que la moyenne empirique

$\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ et la variance empirique $V_n = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ sont des variables aléatoires indépendantes. Le but de cet exercice est de démontrer que la loi de X_i est alors la loi gaussienne $\mathcal{N}(m, \sigma^2)$.

On note ψ la fonction caractéristique de X_i. On suppose $m = 0$.

1. Calculer $\mathbb{E}[(n-1)V_n]$ en fonction de σ^2. Montrer que : pour tout réel t,
$$\mathbb{E}[(n-1)V_n\, \mathrm{e}^{itn\bar{X}_n}] = (n-1)\psi(t)^n\sigma^2.$$
2. En développant V_n dans l'égalité précédente, vérifier que :
$$\mathbb{E}[(n-1)V_n\, \mathrm{e}^{itn\bar{X}_n}] = -(n-1)\psi''(t)\psi(t)^{n-1} - (n-1)\psi'(t)^2\psi(t)^{n-2}.$$
3. En déduire que ψ est solution de l'équation différentielle :
$$\begin{cases} \dfrac{\psi''}{\psi} - (\dfrac{\psi'}{\psi})^2 = -\sigma^2, \\ \psi(0) = 1, \quad \psi'(0) = 0. \end{cases}$$
4. En déduire que la loi des variables X_i est la loi gaussienne $\mathcal{N}(0, \sigma^2)$.
5. Que peut-on dire si l'on ne suppose plus $m = 0$?

△

Deuxième partie

Statistique

VII

Introduction à la statistique : un exemple

Un fabricant de composants électroniques désire avoir des renseignements sur le temps de vie moyen θ des composants de type A. De manière générale, on sait que le temps T d'attente de panne des composants électroniques suit une loi exponentielle de paramètre $\lambda > 0$ inconnu. Cette connaissance peut provenir de considérations théoriques ou empiriques. On a donc $\theta = \mathbb{E}[T] = \lambda^{-1}$.

On aborde avec cet exemple les trois thèmes de la statistique paramétrique que l'on traite dans ce cours :

- **Estimation ponctuelle** : estimation du paramètre inconnu θ, voir le paragraphe VII.1.
- **Test d'hypothèses** : comment tenir compte de l'imprécision de l'estimation pour savoir, par exemple, si $\theta < \theta_*$ ou si $\theta \geq \theta_*$, avec θ_* une valeur de référence fixée, voir le paragraphe VII.2.
- **Intervalle de confiance** : donner la précision de l'estimation, voir le paragraphe VII.3.

VII.1 Estimation ponctuelle

On désire obtenir une estimation du temps de vie moyen θ. Pour cela on considère n composants de type A. Leurs temps de vie $T_1, \ldots, T_n$ sont des variables aléatoires que l'on suppose indépendantes et de loi exponentielle de paramètre $\lambda = \frac{1}{\theta} > 0$ inconnu. On note le vecteur de taille n : $T = T[n] = (T_1, \ldots, T_n)$. La loi forte des grands nombres assure que :

$$\hat{\theta}_n(T) = \frac{1}{n}\sum_{i=1}^{n} T_i \underset{n\to\infty}{\longrightarrow} \mathbb{E}[T_1] = \theta \quad \text{p.s.},$$

$$\text{et} \quad \hat{\theta}'_n(T) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} T_i^2} \underset{n\to\infty}{\longrightarrow} \sqrt{\frac{1}{2}\,\mathbb{E}[T_1^2]} = \theta \quad \text{p.s.}$$

Pour la deuxième convergence, on a également utilisé la continuité de la fonction $t \to \sqrt{t}$. Donc les quantités $\hat{\theta}_n$ et $\hat{\theta}'_n$ permettent de donner une estimation du temps de vie moyen des composants de type A. Lors d'une expérimentation, on observe les réalisations $T_1(\omega) = t_1, \ldots, T_n(\omega) = t_n$. Dans ce cas on estimera θ par $\hat{\theta}_n(t)$ ou $\hat{\theta}'_n(t)$, avec $t = t[n] = (t_1, \ldots, t_n)$. Afin d'alléger les notations, on écrira $\hat{\theta}_n$ pour $\hat{\theta}_n(T)$ et $\hat{\theta}'_n$ pour $\hat{\theta}'_n(T)$. Les estimateurs $\hat{\theta}_n$ et $\hat{\theta}'_n$ convergent p.s. vers la vraie valeur θ, on parle d'estimateurs **convergents**.

Peut-on comparer les estimateurs $\hat{\theta}_n$ et $\hat{\theta}'_n$? Lequel est le meilleur ? Peut-on améliorer un estimateur ? On répondra à ces questions au chapitre VIII.

VII.2 Test d'hypothèses

La garantie du constructeur pour les composants électroniques est de 2 ans. Il peut accepter au plus un taux de 10% de pièces tombant en panne avant 2 ans. Le fabricant désire donc s'assurer que $\mathbb{P}(T_1 \geq 2) \geq 0,9$ soit $\theta \geq 2/\log(1/0,9) = \theta_* \simeq 19$. Il faut donc savoir si l'hypothèse $\theta < \theta_*$ est réaliste, auquel cas il faut par exemple revoir la conception des composants de type A.

Avant de poursuivre le calcul, il faut préciser les priorités du constructeur. En effet les n réalisations des temps de panne ne lui donnent pas la valeur exacte de θ, mais une approximation $\hat{\theta}_n$. À partir de cette approximation, le constructeur doit prendre une décision : mettre en place une nouvelle chaîne de production et diminuer ainsi le taux de défaillance, ou bien ne rien faire et accepter le taux de défaillance actuel. On suppose qu'un taux de défaillance supérieur à 10%, c'est-à-dire $\theta < \theta_*$, met en péril la survie de l'entreprise. Dans ce cas le constructeur devra revoir la conception des composants de type A au moindre soupçon que $\theta < \theta_*$. L'hypothèse cruciale $\{\theta < \theta_*\}$ porte le nom d'**hypothèse nulle**.

L'**erreur de 1ère espèce** consiste à rejeter l'hypothèse nulle alors qu'elle est vraie. C'est ce risque que le constructeur cherche à maîtriser en priorité. Il se fixe donc une probabilité d'erreur, α, maximale, aussi appelée **seuil**. On choisit par exemple 5%. On note $\mathbb{P}_\theta$ la probabilité sous laquelle les variables aléatoires $T_1, \ldots, T_n$ sont indépendantes et de même loi exponentielle de paramètre $1/\theta$. On

suppose qu'il existe z_0 tel que pour tout $\theta \in [0, \theta_*]$, $\mathbb{P}_\theta(\hat{\theta}_n \geq z_0) \leq 5\%$. Ainsi si on observe $\hat{\theta}_n(t) \geq z_0$, il n'est pas raisonnable de supposer que $\theta \in [0, \theta_*]$. En effet cela arrive dans moins de 5% des cas. Si on observe $\hat{\theta}_n(t) \geq z_0$, le fabricant rejettera donc l'hypothèse $\theta < \theta_*$, et il aura raison dans 95% des cas. Il estimera donc, avec une **confiance** de 95%, que le pourcentage de pièces tombant en panne avant 2 ans est inférieur à 10%. En revanche, si le constructeur observe $\hat{\theta}_n(t) < z_0$, alors il est plausible que $\theta < \theta_*$. Dans ce cas, le constructeur accepte l'hypothèse $\{\theta < \theta_*\}$, et décide de revoir la conception des composants de type A.

On continue maintenant les calculs. La loi de $\hat{\theta}_n$ est la loi $\Gamma(n/\theta, n)$. La fonction $h_n(\theta, z) = \mathbb{P}_\theta(\hat{\theta}_n \geq z)$ est définie par :

$$h_n(\theta, z) = \Gamma(n)^{-1} \int_{nz/\theta}^{\infty} x^{n-1} \, \mathrm{e}^{-x} \, dx.$$

Cette fonction est continue, croissante en θ, décroissante en z, et on a $h_n(\theta, 0) = 1$ et $\lim_{z\to\infty} h_n(\theta, z) = 0$. On peut donc trouver $z_0 > 0$ tel que $\mathbb{P}_\theta(\hat{\theta}_n \geq z_0) \leq 5\%$ pour tout $\theta \in [0, \theta_*]$. En fait on résout $h_n(\theta_*, z_0) = 0,05$. On obtient les valeurs approchées suivantes pour z_0 en fonction du nombre d'observations n :

n	1	10	100	1000
z_0	56,9	29,9	22,21	19,98

.

On a bien sûr $z_0 \geq \theta_*$, mais on remarque que z_0 se rapproche de θ_* quand n tend vers l'infini. Cela provient du fait que $\hat{\theta}_n$ converge p.s. vers θ_* sous $\mathbb{P}_{\theta_*}$ quand n tend vers l'infini. En effet, soit $\varepsilon > 0$. Pour tout $\theta \in [0, \theta_*]$, on a par croissance, $\mathbb{P}_\theta(\hat{\theta}_n \geq \theta_* + \varepsilon) \leq \mathbb{P}_{\theta_*}(\hat{\theta}_n \geq \theta_* + \varepsilon)$. Or cette dernière quantité converge vers 0, car $\hat{\theta}_n$ converge vers θ_* $\mathbb{P}_{\theta_*}$-p.s. Donc pour n grand, on a $z_0 \leq \theta_* + \varepsilon$. Comme de plus $z_0 \geq \theta_*$, on en déduit que $\lim_{n\to\infty} z_0 = \theta_*$.

VII.3 Intervalle de confiance

Enfin le fabricant est tenu de fournir à l'entreprise qui utilise son composant un intervalle de confiance sur son estimation du temps de vie de ce composant. Il doit donc fournir des valeurs θ_+ et θ_- telles que la probabilité pour que l'intervalle de confiance (aléatoire) $[\theta_-, \theta_+]$ contienne la vraie valeur θ soit supérieure à 95%. Cette probabilité est appelée le niveau de l'intervalle de confiance. On cherche donc à construire θ_- et θ_+, à partir des observations $t_1, \ldots, t_n$, tels que pour tout $\theta > 0$, on ait :

$$\mathbb{P}_\theta(\theta \in [\theta_-, \theta_+]) \geq 0,95.$$

On rappelle que le vecteur $T = T[n]$ dépend de n. Grâce au théorème central limite, on sait que la suite de v.a. $\left(\sqrt{n}(\hat{\theta}_n - \theta), n \geq 1\right)$ converge en loi vers la loi gaussienne $\mathcal{N}(0, \mathrm{Var}_\theta(T_1))$. Dans le cas particulier de la loi exponentielle, la variance de T_1 s'exprime simplement en fonction de θ : $\mathrm{Var}_\theta(T_1) = \theta^2$. Comme de plus $\hat{\theta}_n$ converge p.s. vers θ, on déduit du théorème de Slutsky V.32, que la suite $\left(\dfrac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\theta}_n}, n \geq 1\right)$ converge en loi vers la loi gaussienne centrée réduite $\mathcal{N}(0,1)$. Par conséquent pour n grand, on a :

$$\mathbb{P}_\theta\left(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \in [-r, r]\right) = \mathbb{P}_\theta\left(\theta \in \left[\hat{\theta}_n - \frac{r\hat{\theta}_n}{\sqrt{n}}, \hat{\theta}_n + \frac{r\hat{\theta}_n}{\sqrt{n}}\right]\right)$$
$$\simeq \int_{[-r,+r]} \mathrm{e}^{-x^2/2} \ \frac{dx}{\sqrt{2\pi}}. \qquad \text{(VII.1)}$$

Le membre de droite est approximativement égal à 95% si on choisit $r_0 \simeq 1,96$ (cf. la table des quantiles de la loi $\mathcal{N}(0,1)$ au paragraphe XI.1). Le fabricant, s'il a un grand nombre de résultats $t[n] = (t_1, \ldots, t_n)$, assurera avec une probabilité asymptotique de 95% que l'intervalle **aléatoire** $\left[\hat{\theta}_n(t[n]) \pm \dfrac{r_0\hat{\theta}_n(t[n])}{\sqrt{n}}\right]$ contient le temps de vie moyen des composants de type A. Il s'agit dans cet exemple d'un intervalle de confiance dont le niveau est de 95% quand le nombre d'expériences est grand. On parle d'intervalle de confiance de niveau asymptotique 95%. On remarque enfin que l'on a peu de contrôle sur l'estimation (VII.1). Un ordre de grandeur peut être donné par l'inégalité de Berry-Esséen. Dans cet exemple précis la majoration dans l'inégalité de Berry-Esséen est indépendante de θ ; elle est de l'ordre d'une constante divisée par $\sqrt{n}$. De manière générale, pour effectuer une approche asymptotique, il est nécessaire de disposer d'un grand nombre de résultats.

Dans notre cas particulier, on peut également donner des intervalles de confiance de niveau exact. En effet la variable aléatoire $\hat{\theta}_n$ suit une loi $\Gamma(n/\theta, n)$. On en déduit que $n\hat{\theta}_n/\theta$ suit une loi $\Gamma(1, n)$. Soit $a_n < b_n$ tels que :

$$\Gamma(n)^{-1} \int_{a_n}^{b_n} x^{n-1}\,\mathrm{e}^{-x} \ dx = 0,95.$$

On a alors $\mathbb{P}_\theta\left(n\hat{\theta}_1(T[n])/\theta \in [a_n, b_n]\right) = 0,95$. On en déduit donc que :

$$\mathbb{P}_\theta\left(\theta \in \left[\frac{n\hat{\theta}_n}{b_n}, \frac{n\hat{\theta}_n}{a_n}\right]\right) = 0,95.$$

Il existe bien sûr plusieurs choix pour a_n et b_n : on peut désirer un intervalle symétrique autour de $\hat{\theta}_n$, ou bien un intervalle de longueur minimale,...

Enfin à l'aide de ce calcul exact, on peut vérifier la validité de l'approche asymptotique. Par exemple, en choisissant $n/a_n = (1 - r_0/\sqrt{n})$ et $n/b_n = (1 + r_0/\sqrt{n})$, on calcule numériquement le niveau de l'intervalle de confiance :

$$\alpha_n = \mathbb{P}_\theta\left(\theta \in \left[\hat{\theta}_1(T[n]) \pm \frac{r_0\hat{\theta}_1(T[n])}{\sqrt{n}}\right]\right) \quad \text{où} \quad r_0 = 1,96.$$

On obtient les valeurs numériques suivantes (indépendantes de θ) :

n	1	5	10	100	1000
α_n	0,71	0,87	0,90	0,94	0,95

.

On retrouve bien que le niveau exact α_n est proche du niveau asymptotique 95% pour n grand. En revanche pour les petites valeurs de n, l'intervalle de confiance a, dans cet exemple, un niveau exact nettement plus faible que le niveau asymptotique. Il faut savoir manipuler les intervalles de confiance de niveau asymptotique avec précaution.

Enfin, on compare $\alpha_n - 0,95$ et la majoration Berry-Esséen du théorème V.33 :

$$\beta_n = \left|\mathbb{P}_\theta\left(\theta \in \left[\hat{\theta}_n \pm \frac{r_0\theta}{\sqrt{n}}\right]\right) - \int_{-r_0}^{r_0} \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}\right| \leq 2C\frac{\mathbb{E}_\theta[|T_1-\theta|^3]}{\mathrm{Var}_\theta(T_1)^{3/2}\sqrt{n}},$$

où $C \leq 0,8$ et $r_0 = 1,96$. Le membre de droite est donc majoré par $2 * 0,8(12\,\mathrm{e}^{-1} - 2)/\sqrt{n} = c_0/\sqrt{n}$. On remarque que l'on utilise la vraie valeur de la variance dans le calcul de β_n, mais de sa valeur approchée pour α_n (voir aussi le théorème V.34). Le tableau suivant est indépendant des valeurs de θ :

n	1	5	10	100
$\lvert\alpha_n - 0,95\rvert$	0,24	0,081	0,046	0,005
β_n	0,018	0,006	0,005	0,0006
$c_0/\sqrt{n}$	1,93	0,86	0,61	0,19

.

On remarque que l'incertitude sur la variance induit une erreur plus grande sur la validité de l'intervalle de confiance ($|\alpha_n - 0,95| \geq \beta_n$). On voit aussi sur cet exemple, que la majoration de Berry-Esséen ne donne pas du tout le bon ordre de grandeur de la convergence pour n grand.

VIII

Estimation ponctuelle

On présente dans ce chapitre une introduction à l'estimation dans un modèle paramétrique (pour plus de détails, le lecteur pourra consulter les ouvrages [10, 11], [2] ou [3]).

Le paragraphe VIII.1 est consacré aux hypothèses sur le modèle paramétrique. On donne dans le paragraphe VIII.2 les définitions et quelques propriétés des estimateurs des paramètres du modèle. On étudie dans le paragraphe VIII.3 les estimateurs de substitution, les estimateurs des moments, et l'estimateur du maximum de vraisemblance. Ce dernier possède d'excellentes propriétés, mais on ne peut pas toujours le calculer explicitement. Dans le paragraphe VIII.4 on considère le risque quadratique pour comparer des estimateurs (à horizon fini). En particulier, on traite en détail le modèle gaussien au paragraphe VIII.4.4. La notion de statistique exhaustive introduite dans le paragraphe VIII.5 permet de résumer les données observées et le théorème de Rao-Blackwell assure que l'on peut améliorer un estimateur en utilisant une statistique exhaustive. Dans de nombreux cas, on dispose d'échantillons de grande taille. On peut alors utiliser une approche asymptotique (horizon infini). Cette approche est développée au paragraphe VIII.6. (En général, elle donne également de bonnes intuitions pour l'étude à horizon fini.) Sous des conditions assez générales, l'estimateur du maximum de vraisemblance a un excellent comportement asymptotique.

VIII.1 Hypothèses sur le modèle

On considère essentiellement les modèles d'**échantillonnage**. Cela correspond à l'observation d'un échantillon de n variables aléatoires $X_1, \ldots, X_n$ à valeurs dans $\mathcal{X}$ ($\mathcal{X} = \mathbb{R}$ ou $\mathcal{X} = \mathbb{R}^d$) indépendantes et de même loi P. Par convention la loi de

$X = X[n] = (X_1, \dots, X_n)$ est notée $\mathrm{P}^{\otimes n}$. Il s'agit de la loi produit. On suppose que la loi inconnue P appartient à une famille de probabilités $\mathcal{P} = \{\mathrm{P}_\theta, \theta \in \Theta\}$, où θ est un paramètre et Θ un sous-ensemble mesurable de $\mathbb{R}^p$. On dit que le modèle est **paramétrique** et que la dimension du paramètre est p. Le modèle est **identifiable** si $\theta_1 \neq \theta_2$ implique $\mathrm{P}_{\theta_1} \neq \mathrm{P}_{\theta_2}$. Par exemple les familles $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$, où $\mathcal{E}(\lambda)$ désigne la loi exponentielle de paramètre λ, et $\mathcal{P} = \left\{\mathcal{N}(m, \sigma^2); m \in \mathbb{R}, \sigma > 0\right\}$ correspondent à des modèles paramétriques identifiables. Dans le deuxième cas le paramètre $\theta = (m, \sigma)$ est de dimension 2.

Le but de l'estimation paramétrique est d'identifier la probabilité inconnue P_θ ou le paramètre inconnu θ à partir de la réalisation d'un échantillon.

On étudie essentiellement des modèles d'échantillonnage où la famille de probabilité $\mathcal{P}$ est constituée **soit** de lois possédant une densité $f_\theta(x) = p(x;\theta)$ **soit** de lois discrètes $\mathrm{P}_\theta(X = x) = p(x;\theta)$. Par abus de langage, on appelle dans tous les cas $p(\cdot;\theta)$ la **densité** de la loi P_θ. (C'est effectivement la densité de la loi par rapport à une mesure de référence. Dans le cas de v.a. continues, la mesure de référence est la mesure de Lebesgue. Dans le cas de v.a. discrètes à valeurs dans E, la mesure de référence est la mesure de comptage sur E.) Enfin on note $\Delta_\theta = \{x; p(x;\theta) > 0\} \subset \mathcal{X}$ le support de la loi de P_θ. On le note Δ s'il est indépendant de $\theta \in \Theta$.

Comme les variables aléatoires $X_1, \dots, X_n$ sont indépendantes, la densité de l'échantillon $X = (X_1, \dots, X_n)$ est la densité produit :

$$p_n(x;\theta) = p(x_1;\theta) \cdots p(x_n;\theta), \quad \text{où} \quad x = (x_1, \dots, x_n) \in \mathcal{X}^n.$$

Dans tout ce qui suit, les énoncés et les démonstrations concernent le cas où $\mathcal{P}$ est une famille de lois continues. Ils se généralisent au cas où $\mathcal{P}$ est une famille de lois discrètes en remplaçant le signe $\int dx$ par $\sum_x$. Cette distinction artificielle est due au fait que l'on n'utilise pas la théorie de la mesure.

Exemple. On note $\mathcal{B}(1,\theta)$ la loi de Bernoulli de paramètre $\theta \in [0,1]$. Il s'agit en fait d'une loi binomiale de paramètre $(1,\theta)$. La famille $\mathcal{P} = \{\mathcal{B}(1,\theta); \theta \in \Theta = [0,1]\}$ définit un modèle paramétrique identifiable discret. ◊

Enfin on notera $\mathbb{E}_\theta[f(X_1, \dots, X_n)]$ l'espérance de $f(X_1, \dots, X_n)$ où le vecteur $X = (X_1, \dots, X_n)$ a pour loi $\mathrm{P}_\theta^{\otimes n}$. Ainsi dans l'exemple ci-dessus on a $\mathbb{E}_\theta[(X_1 + \cdots + X_n)/n] = \mathbb{E}_\theta[X_1] = \theta$.

VIII.2 Statistiques et estimateurs

La définition suivante assure qu'une statistique peut être calculée à partir des observations.

Définition VIII.1. *Une* ***statistique*** S *est une fonction de l'échantillon* $X = (X_1, \ldots, X_n)$ *réelle ou vectorielle. Cette fonction est indépendante de* $\mathrm{P} \in \mathcal{P}$.

Exemple. Les fonctions $S(X) = \sum_{i=1}^n X_i$ et $S(X) = X_1$ sont des statistiques de l'échantillon $(X_1, \ldots, X_n)$.

Si on considère le modèle gaussien à moyenne μ_0 fixée et variance inconnue, $\mathcal{P} = \{\mathcal{N}(\mu_0, \sigma^2), \sigma > 0\}$, alors la variable aléatoire $\frac{1}{n}\sum_{k=1}^n (X_k - \mu_0)^2$ est une statistique de l'échantillon $(X_1, \ldots, X_n)$; elle converge p.s. vers le paramètre σ^2. En revanche, si on considère le modèle gaussien à moyenne et variance inconnues, $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0\}$, alors la variable aléatoire $\frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2$ n'est plus une statistique car elle dépend du paramètre (inconnu) μ. $\Diamond$

On a vu dans le chapitre VII que le premier objectif est d'estimer le paramètre θ. Plus généralement on peut chercher à estimer une fonction $g(\theta) \in \mathbb{R}^k$ de ce paramètre.

Définition VIII.2. *Un* ***estimateur*** δ *de* $g(\theta)$ *est une statistique à valeurs dans* $g(\Theta)$.

Exemple. On considère un échantillon de taille n correspondant à un modèle de Bernoulli $\mathcal{P} = \{\mathcal{B}(1, \theta); \theta \in [0, 1]\}$. Ainsi $\delta = \frac{1}{n}\sum_{i=1}^n X_i$ et $\delta = 1/2$ sont des estimateurs de θ, mais ils ont des comportements asymptotiques très différents. $\Diamond$

Définition VIII.3. *Une suite d'estimateurs* $(\delta_n, n \in \mathbb{N}^*)$ *de* $g(\theta)$ *où* δ_n *est une fonction de l'échantillon* $X_1, \ldots, X_n$ *de taille* n*, est* ***convergente*** *si pour tout* $\theta \in \Theta$*, on a :*

$$\boxed{\lim_{n\to\infty} \delta_n = g(\theta) \quad \mathbb{P}_\theta\text{-p.s.}}$$

Par brièveté, on confondra estimateur et suite d'estimateurs. On utilise souvent l'anglicisme "consistant" pour convergent. Dans la littérature, on dit qu'un estimateur est faiblement convergent s'il converge en probabilité : pour tout $\theta \in \Theta$, on a :

$$\lim_{n\to\infty} \mathbb{P}_\theta(|\delta_n - g(\theta)| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

Par souci de précision, si l'estimateur converge presque sûrement, on dit parfois qu'il est "fortement convergent".

Exemple VIII.4. Soit le modèle d'échantillonnage exponentiel $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$. Les estimateurs :

$$\hat{\lambda}_n^{(1)} = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i} \quad \text{et} \quad \hat{\lambda}_n^{(2)} = \frac{1}{\sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}}$$

sont deux estimateurs convergents de λ. En effet, cela découle de la loi forte des grands nombres avec $\mathbb{E}_\lambda[X_1] = \lambda^{-1}$ et $\mathbb{E}_\lambda[X_1^2] = 2\lambda^{-2}$ ainsi que du théorème de continuité pour la convergence p.s. (lemme V.2). ◊

VIII.3 Construction d'estimateurs convergents

On donne trois méthodes de construction d'estimateurs convergents.

VIII.3.1 Méthode de substitution

On suppose que l'on dispose d'un estimateur $\hat{\theta}_n$ convergent de θ. On peut alors construire un estimateur de $g(\theta)$ en substituant θ par $\hat{\theta}_n$. On suppose que $g : \Theta \to g(\Theta)$ est continue, alors l'estimateur $\hat{g}_n = g(\hat{\theta}_n)$ est un estimateur convergent de $g(\theta)$. C'est une conséquence du théorème de continuité pour la convergence p.s. (lemme V.2). On a utilisé cette méthode dans l'exemple VIII.4 ci-dessus.

VIII.3.2 Méthode des moments

Exemple. On considère le modèle de la loi béta $\mathcal{P} = \{\beta(a,b); a > 0, b > 0\}$. Le paramètre $\theta = (a,b)$ est de dimension 2. On désire estimer θ. Soit X_1 de loi $\beta(a,b)$, on a :

$$\mathbb{E}_{(a,b)}[X_1] = \frac{a}{a+b} = c \quad \text{et} \quad \mathbb{E}_{(a,b)}[X_1(1-X_1)] = \frac{ab}{(a+b)(a+b+1)} = d.$$

En inversant le système, on obtient $a = \dfrac{cd}{c-d-c^2}$ et $b = \dfrac{(1-c)d}{c-d-c^2}$. On construit à l'aide de la loi forte des grands nombres des estimateurs convergents de c et d :

$$\hat{c}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{et} \quad \hat{d}_n = \frac{1}{n}\sum_{i=1}^n X_i(1-X_i),$$

où les v.a. $X_1, \ldots, X_n$ sont indépendantes de loi $\beta(a, b)$. On déduit de la méthode de substitution que :

$$\hat{a}_n = \frac{\hat{c}_n \hat{d}_n}{\hat{c}_n - \hat{d}_n - \hat{c}_n^2} \quad \text{et} \quad \hat{b}_n = \frac{(1 - \hat{c}_n)\hat{d}_n}{\hat{c}_n - \hat{d}_n - \hat{c}_n^2}$$

sont des estimateurs convergents de a et b. ◊

La méthode des moments consiste à trouver une fonction m, inversible et continue, et une fonction mesurable φ, telles que $\mathbb{E}_\theta[|\varphi(X_1)|] < \infty$ et $m(\theta) = \mathbb{E}_\theta[\varphi(X_1)]$ pour tout $\theta \in \Theta$. Un **estimateur des moments** de θ est alors :

$$\boxed{\hat{\theta}_n = m^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)\right).}$$

Les estimateurs des moments sont convergents.

VIII.3.3 Le maximum de vraisemblance

Exemple. Au cours d'une enquête, on recherche un suspect dont la taille est d'environ 1m80. Afin d'orienter rapidement l'enquête, doit-on rechercher plutôt un suspect masculin ou féminin ?

On peut modéliser les données du "National Center for Health Statistics" (USA) sur la période 1988-1994 pour la taille d'un homme par une loi gaussienne de moyenne m_H =1m76 et d'écart type σ_H =0m073, et la taille d'une femme par une loi gaussienne de moyenne m_F =1m62 et d'écart type σ_F =0m069. La densité de la loi gaussienne de moyenne m et d'écart type σ est :

$$p(x; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x-m)^2/2\sigma^2}, \quad x \in \mathbb{R}.$$

Si la taille du suspect est 1m80, alors on se concentre sur la recherche d'un suspect masculin. Le choix est en fait guidé par l'allure de la densité, voir la figure VIII.1. Et plus particulièrement ayant observé la taille x_0 =1m80, il est raisonnable de supposer que le suspect est un homme car $p(x_0; \theta_H) \geq p(x_0; \theta_F)$, où $\theta_H = (m_H, \sigma_H)$ et $\theta_F = (m_F, \sigma_F)$. On a choisi le paramètre $\theta \in \Theta = \{\theta_F, \theta_H\}$ qui maximise la fonction $\theta \mapsto p(x_0; \theta)$. ◊

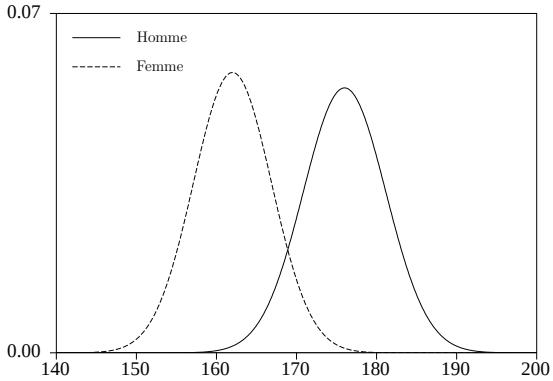


Figure VIII.1. Densités gaussiennes pour la taille d'un homme et d'une femme.

Soit $p(x_1;\theta)$ la densité de la loi de X_1 et $p_n(x;\theta) = p(x_1;\theta)\cdots p(x_n;\theta)$ la densité de la loi de l'échantillon $(X_1,\ldots,X_n)$ où $x = (x_1,\ldots,x_n) \in \mathcal{X}^n$.

Définition VIII.5. *La fonction réelle $\theta \mapsto p_n(x;\theta)$ à $x \in \mathcal{X}^n$ fixé est appelée la* ***vraisemblance*** *de l'échantillon ou de la réalisation $x = (x_1,\ldots,x_n)$. Elle est définie sur Θ. On suppose que pour tout $x \in \mathcal{X}^n$, il existe* ***une et une seule valeur*** *de $\theta \in \Theta$ telle que la vraisemblance soit maximale. On la note $\hat{\theta}_n(x_1,\ldots,x_n)$. La variable aléatoire $\hat{\theta}_n = \hat{\theta}_n(X_1,\ldots,X_n)$, à valeurs dans Θ, est appelée* ***estimateur du maximum de vraisemblance*** *(EMV) de θ.*

On définit également la **log-vraisemblance** du modèle comme le logarithme de la vraisemblance ; c'est la fonction :

$$\boxed{\theta \mapsto L_n(x;\theta) = \log(p_n(x;\theta)) = \sum_{i=1}^{n} \log(p(x_i;\theta)).}$$

Maximiser la vraisemblance ou la log-vraisemblance revient au même. Maximiser la log-vraisemblance donne lieu parfois à des calculs plus simples.

Exemple. On considère un modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu,\sigma_0^2);\mu \in \mathbb{R}\}$, où la variance est connue. La log-vraisemblance s'écrit :

$$L_n(x;\mu) = -\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma_0^2} - \frac{n}{2}\log(2\pi\sigma_0^2).$$

Comme on a $L_n(x;\pm\infty) = -\infty$, on en déduit que le maximum de $L_n(x;\cdot)$ est atteint pour μ tel que $\dfrac{\partial L_n}{\partial \mu}(x;\mu) = 0$. On obtient $-2n\mu + 2\sum_{i=1}^n x_i = 0$. Le maximum de la vraisemblance est atteint pour $\hat{\mu}_n(x) = \frac{1}{n}\sum_{i=1}^n x_i$. L'estimateur du maximum de vraisemblance est donc $\hat{\mu}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$. On retrouve dans le cas gaussien la moyenne empirique. Cet estimateur est convergent par la loi forte des grands nombres. ◊

Exemple VIII.6. On considère un modèle uniforme $\mathcal{P} = \left\{U_{[0,\theta]}; \theta > 0\right\}$. La vraisemblance est :

$$p_n(x;\theta) = \mathbf{1}_{\{\min_{1\le i\le n} x_i \ge 0\}} \mathbf{1}_{\{\max_{1\le i\le n} x_i \le \theta\}}\, \theta^{-n}.$$

La vraisemblance est maximale pour $\hat{\theta}_n(x) = \max_{1\le i\le n} x_i$ (voir la figure VIII.2). L'estimateur du maximum de vraisemblance est $\hat{\theta}_n = \max_{1\le i\le n} X_i$.

Cet estimateur est convergent. En effet la suite de v.a. $(\hat{\theta}_n, n \in \mathbb{N}^*)$ est croissante. Elle est donc convergente p.s. dans $]0,+\infty]$. On note M sa limite. On remarque ensuite que pour tout $\varepsilon > 0$, on a :

$$\mathbb{P}_\theta\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = \mathbb{P}_\theta(\hat{\theta}_n \le \theta - \varepsilon) = \mathbb{P}_\theta(X_1 \le \theta - \varepsilon)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n \mathbf{1}_{\{\varepsilon \le \theta\}}.$$

On en déduit que l'estimateur $\hat{\theta}_n$ converge en probabilité vers θ. Comme la convergence p.s. implique la convergence en probabilité, on en déduit que $M = \theta$ $\mathbb{P}_\theta$-p.s. L'estimateur du maximum de vraisemblance est donc convergent. En remarquant que $\mathbb{E}_\theta[X_1] = \theta/2$, on peut également construire un estimateur convergent de θ par la méthode des moments : $\hat{\theta}'_n = \frac{2}{n}\sum_{i=1}^n X_i$. ◊

Proposition VIII.7. *On suppose que $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ associé à la famille de loi $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ existe. On suppose que la fonction $g : \Theta \to g(\Theta)$ est bijective et mesurable. Alors $g(\hat{\theta}_n)$ est l'estimateur du maximum de vraisemblance de $g(\theta)$ associé à la famille paramétrique $\mathcal{P}_g = \left\{\mathrm{Q}_\beta = \mathrm{P}_{g^{-1}(\beta)}; \beta \in g(\Theta)\right\}$. On dit que $g(\hat{\theta}_n)$ est l'estimateur du maximum de vraisemblance de $g(\theta)$.*

La démonstration est immédiate. Par convention, on dit que $g(\hat{\theta}_n)$ est l'estimateur du maximum de vraisemblance, même si g n'est pas bijective.

Exemple. Soit un modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$. La log-vraisemblance est :

$$L_n(x;\lambda) = -\lambda \sum_{i=1}^n x_i + n \log \lambda.$$

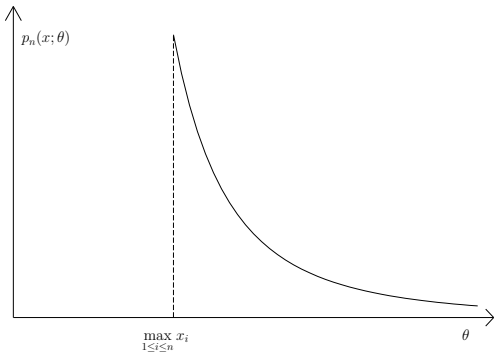


Figure VIII.2. La vraisemblance d'un échantillon suivant la loi uniforme.

L'estimateur du maximum de vraisemblance de λ est $\hat{\lambda}_n = \dfrac{n}{\sum_{i=1}^n X_i}$, et l'estimateur du maximum de vraisemblance de $\theta = \lambda^{-1}$ est $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$. $\Diamond$

On définit, quand elle existe, l'**entropie** de la loi P_θ par :

$$\boxed{H_\theta = -\mathbb{E}_\theta[\log(p(X_1;\theta))]}$$

(ainsi si X_1 est une v.a. continue alors $H_\theta = -\int p(x;\theta)\log p(x;\theta)\,dx$, et si X_1 est une v.a. discrète alors $H_\theta = -\sum_{x\in\Delta_\theta} p(x;\theta)\log p(x;\theta)$). On retrouve cette quantité au signe près en physique statistique et en théorie de l'information. L'entropie mesure le désordre d'un système de particules ou la quantité d'information d'un système.

On considère l'application, quand elle existe, $\mathcal{H}_{\theta_0}$ définie sur Θ par :

$$\mathcal{H}_{\theta_0} :\ \theta \mapsto \mathcal{H}_{\theta_0}(\theta) = \mathbb{E}_{\theta_0}\left[\log p(X_1;\theta)\right].$$

On remarque que $\mathcal{H}_{\theta_0}(\theta_0) = -H_{\theta_0}$.

Lemme VIII.8. *On suppose que $\mathbb{E}_{\theta_0}[|\log p(X_1;\theta)|]$ est fini pour tout $\theta, \theta_0 \in \Theta$. Pour tout $\theta_0 \in \Theta$, l'application $\mathcal{H}_{\theta_0}$ existe et atteint son maximum pour $\theta = \theta_0$. De plus le maximum est unique pour tout $\theta_0 \in \Theta$ si et seulement si le modèle est identifiable.*

Démonstration. On a $\log z \leq z - 1$ pour $z \geq 0$. Donc pour $\theta \in \Theta$, il vient (dans le cas des v.a.c.) :

$$\begin{aligned}\mathcal{H}_{\theta_0}(\theta) - \mathcal{H}_{\theta_0}(\theta_0) &= \int p(x;\theta_0) \log p(x;\theta)\, dx - \int p(x;\theta_0) \log p(x;\theta_0)\, dx \\ &= \int p(x;\theta_0) \log\left(\frac{p(x;\theta)}{p(x;\theta_0)}\right) dx \\ &\leq \int p(x;\theta_0) \left(\frac{p(x;\theta)}{p(x;\theta_0)} - 1\right) dx \\ &= 0.\end{aligned}$$

On en déduit que $\mathcal{H}_{\theta_0}(\theta) \leq \mathcal{H}_{\theta_0}(\theta_0)$. Enfin l'égalité n'a lieu que si $\log\left(\frac{p(x;\theta)}{p(x;\theta_0)}\right) = \frac{p(x;\theta)}{p(x;\theta_0)} - 1$ pour presque tout $x \in \Delta_{\theta_0}$. Or $\log z = z - 1$ implique $z = 1$. Donc si on a l'égalité $\mathcal{H}_{\theta_0}(\theta) = \mathcal{H}_{\theta_0}(\theta_0)$, alors $p(x;\theta) = p(x;\theta_0)$ pour presque tout $x \in \Delta_{\theta_0}$. Ce qui entraîne que $\theta = \theta_0$ pour tout $\theta_0 \in \Theta$ si et seulement si le modèle est identifiable. □

Le théorème suivant assure que sous certaines hypothèses, l'estimateur du maximum de vraisemblance est convergent.

Théorème VIII.9. *On fait les hypothèses suivantes :*

1. *Le modèle est identifiable, l'ensemble Θ est compact et les supports Δ_θ sont indépendants de θ.*
2. *La fonction $(x, \theta) \mapsto p(x;\theta)$ est bornée.*
3. *La fonction $\theta \mapsto p(x;\theta)$ est continue.*
4. *L'entropie H_θ est bien définie pour tout $\theta \in \Theta$ (cela revient à dire que $\log p(X_1, \theta)$ est $\mathbb{P}_\theta$ intégrable).*

Sous ces hypothèses, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est un estimateur de θ convergent.

Démonstration. On donne seulement les idées de la démonstration. L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est tel que :

$$\frac{1}{n}\sum_{i=1}^{n} \log p(X_i;\hat{\theta}_n) = \max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log p(X_i;\theta).$$

La loi forte des grands nombres assure que $\left(\frac{1}{n}\sum_{i=1}^{n}\log p(X_i;\theta), n\in\mathbb{N}^*\right)$ converge $\mathbb{P}_{\theta_0}$-p.s. vers $\mathbb{E}_{\theta_0}[\log p(X_1;\theta)] = \mathcal{H}_{\theta_0}(\theta)$, et ce pour tout $\theta\in\Theta$. On admet que sous les hypothèses du théorème, la suite $(\hat{\theta}_n, n\in\mathbb{N}^*)$, converge $\mathbb{P}_{\theta_0}$-p.s. vers θ^* qui satisfait l'égalité ci-dessus à la limite :

$$\mathcal{H}_{\theta_0}(\theta^*) = \max_{\theta\in\Theta}\mathcal{H}_{\theta_0}(\theta).$$

Or d'après le lemme VIII.8, θ^* est unique et égal à θ_0. L'estimateur du maximum de vraisemblance est donc convergent. □

Remarque. Pour montrer la convergence du maximum de vraisemblance, on impose des conditions de régularité sur la densité. Or on verra au paragraphe VIII.6 sur l'exemple du modèle uniforme (voir l'exemple VIII.6 et l'exercice VIII.43) qui ne satisfait pas les hypothèses du théorème, que les convergences sont d'autant meilleures que les densités sont irrégulières. ◊

Exemple. On considère le modèle d'échantillonnage $(X_1,\dots,X_n)$ associé à la loi de Cauchy décentrée de densité : $p(x;\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$, $\theta\in\Theta=\mathbb{R}$. Comme les v.a. $X_1,\dots,X_n$ ne sont pas intégrables, l'estimateur de la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$ n'est pas convergent. En fait, on peut montrer, voir les propos précédents la proposition V.26, que $\limsup_{n\to\infty}\bar{X}_n = +\infty$ et $\liminf_{n\to\infty}\bar{X}_n = -\infty$. L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ annule la dérivée de la vraisemblance. En particulier, il est solution de $\displaystyle\sum_{i=1}^{n}\frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = 0$. Malheureusement, il est difficile de trouver numériquement ces solutions. On peut montrer que la médiane empirique $\hat{\theta}_n$, définie comme la médiane de $\{X_1,\dots,X_n\}$, est un estimateur convergent de θ. ◊

VIII.4 Choix d'un estimateur

On choisit un critère de comparaison, le risque quadratique, entre estimateurs au paragraphe VIII.4.1. On introduit au paragraphe VIII.4.2 l'information de Fisher quand le modèle est régulier. Cette quantité permet de donner une borne inférieure au risque quadratique en absence de biais, voir le paragraphe VIII.4.3. Le modèle gaussien est étudié en détail au paragraphe VIII.4.4.

VIII.4.1 Risque quadratique et comparaison d'estimateurs

On se donne un modèle d'échantillonnage de taille n **fixée**. Les variables aléatoires $X_1,\dots,X_n$ à valeurs dans $\mathcal{X}=\mathbb{R}^d$ sont indépendantes et de même loi. Soit

δ un estimateur de $g(\theta)$. On supposera dans ce paragraphe que l'estimateur δ est de **carré intégrable** : pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[|\delta|^2] < \infty$.

Pour mesurer l'erreur entre l'estimateur et le paramètre, on utilise souvent (mais pas exclusivement) la **fonction de risque quadratique** définie en dimension 1 par :

$$R(\delta, \theta) = \mathbb{E}_\theta \left[(\delta - g(\theta))^2 \right].$$

Il est naturel de chercher les estimateurs qui minimisent la fonction de risque quadratique.

En dimension supérieure, la fonction de risque quadratique est définie par la matrice symétrique positive :

$$R(\delta, \theta) = \mathbb{E}_\theta \left[(\delta - g(\theta))(\delta - g(\theta))^t \right],$$

où par convention, si $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, alors $xx^t = (x_i x_j, 1 \leq i, j \leq p)$ est une matrice de taille $p \times p$. Pour comparer les estimateurs, on utilise alors l'ordre partiel sur les matrices symétriques positives : $A \geq B$ signifie que la matrice $A - B$ est positive.

Définition VIII.10. *On dit que δ_1 est un estimateur **préférable** à δ_2 si : pour tout $\theta \in \Theta$,*

$$R(\delta_1, \theta) \leq R(\delta_2, \theta) \quad \textit{pour tout} \quad \theta \in \Theta.$$

*Si de plus $R(\delta_1, \theta) \neq R(\delta_2, \theta)$ pour au moins une valeur de $\theta \in \Theta$, on dit alors que l'estimateur δ_2 est **inadmissible**.*

Sauf dans des cas triviaux il n'existe pas d'estimateur préférable à tous les autres, voir la remarque qui suit.

Remarque. On suppose pour simplifier que le paramètre peut prendre 2 valeurs : $\Theta = \{\theta_1, \theta_2\}$ et qu'il existe un estimateur $\delta = \delta(X_1)$ de θ préférable à tous les autres estimateurs. Entre autre, δ est préférable aux constantes θ_1 et θ_2. Donc on a :

$$R(\delta, \theta_i) \leq R(\theta_i, \theta_i) = 0, \quad i \in \{1, 2\}.$$

Cela implique que $\delta = \theta_i$ $\mathbb{P}_{\theta_i}$-p.s. pour $i \in \{1, 2\}$. Donc il existe A_1 et A_2 disjoints tels que $\delta = \theta_1$ sur A_1 et $\delta = \theta_2$ sur A_2. De plus on a $\mathbb{P}_{\theta_1}(X_1 \in \delta^{-1}(A_1)) = 1$ et $\mathbb{P}_{\theta_2}(X_1 \in \delta^{-1}(A_2)) = 1$. Les supports des lois P_{θ_1} et P_{θ_2} sont donc disjoints. Cela correspond aux cas triviaux de l'estimation ponctuelle. $\Diamond$

Il est toutefois possible d'obtenir des estimateurs préférables à tous les autres au sein d'une famille restreinte d'estimateurs. On remarque que, en dimension 1 :

$$R(\delta,\theta) = \mathrm{Var}_\theta(\delta) + (\mathbb{E}_\theta[\delta] - g(\theta))^2$$

et que l'on souhaite minimiser $R(\delta,\theta)$ pour tout $\theta \in \Theta$. Il apparaît naturel de considérer les estimateurs tels que $\mathbb{E}_\theta[\delta] = g(\theta)$, puis de chercher à minimiser $\mathrm{Var}_\theta(\delta)$. On peut toutefois obtenir par cette méthode des estimateurs inadmissibles, comme on le verra plus tard.

Définition VIII.11. *Un estimateur δ de $g(\theta)$ est intégrable si $\mathbb{E}_\theta[|\delta|] < +\infty$ pour tout $\theta \in \Theta$. Son **biais** est $\mathbb{E}_\theta[\delta] - g(\theta)$.*

*Un estimateur δ est un estimateur **sans biais** de $g(\theta)$, s'il est intégrable et si $\mathbb{E}_\theta[\delta] = g(\theta)$ pour tout $\theta \in \Theta$.*

Si l'estimateur δ est un estimateur sans biais de $g(\theta)$ et est de carré intégrable, alors on a $R(\delta,\theta) = \mathrm{Var}_\theta(\delta)$.

Exemple. Si X_1 est intégrable pour tout $\theta \in \Theta$, alors la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais de $\mathbb{E}_\theta[X_1]$. ◊

Exemple VIII.12. Pour le modèle de loi uniforme $\mathcal{P} = \{U_{[0,\theta]}; \theta > 0\}$, on a vu dans l'exemple VIII.6 que $\hat{\theta}_n = \max_{1\le i\le n} X_i$ était un estimateur convergent de θ. En revanche cet estimateur est biaisé. En effet, pour tout $x \in \mathbb{R}$, on a par indépendance :

$$\mathbb{P}_\theta(\hat{\theta}_n \le x) = \mathbb{P}_\theta(\forall i \in \{1,\dots,n\}; X_i \le x) = \mathbb{P}_\theta(X_1 \le x)^n.$$

On en déduit que la loi de $\hat{\theta}_n$ possède la densité $\frac{n}{\theta^n} t^{n-1} \mathbf{1}_{[0,\theta]}(t)$. On a donc :

$$\mathbb{E}_\theta[\hat{\theta}_n] = \frac{n}{n+1}\theta.$$

L'estimateur $\hat{\theta}''_n = \frac{n+1}{n}\hat{\theta}_n$ est un estimateur convergent et sans biais de θ. On remarque toutefois que l'estimateur $\hat{\theta}_n$ possède la propriété intéressante de fournir un minorant absolu de θ car p.s. $\hat{\theta}_n \le \theta$. ◊

Exercice VIII.1.
Montrer que les estimateurs $\hat{\theta}_n$ et $\hat{\theta}''_n$ de l'exemple VIII.12 sont des estimateurs inadmissibles. Pour cela on regardera la famille d'estimateurs de la forme $c_n\hat{\theta}_n$. △

Remarque. L'estimateur du maximum de vraisemblance possède en général un biais. En revanche l'absence de biais n'est pas une propriété invariante par transformation : Si δ est un estimateur sans biais de θ, alors $g(\delta)$ est en général un estimateur biaisé de $g(\theta)$. ◊

VIII.4.2 Score, information de Fisher, modèle régulier

Dans le cas où le paramètre θ est vectoriel : $\theta = (\theta_1, \ldots, \theta_p)$, on note $\dfrac{\partial p}{\partial \theta} = \left(\dfrac{\partial p}{\partial \theta_1}, \ldots, \dfrac{\partial p}{\partial \theta_p}\right)$ et $\dfrac{\partial^2 p}{\partial \theta^2} = \left(\dfrac{\partial^2 p}{\partial \theta_i \partial \theta_j}; 1 \leq i, j \leq p\right)$ les dérivées et dérivées secondes de la vraisemblance quand elles existent.

On fait les hypothèses suivantes.

1. Le support des lois P_θ : $\Delta = \{x \in \mathcal{X}; p(x;\theta) > 0\}$ est indépendant de $\theta \in \Theta$.
2. La fonction de vraisemblance est deux fois différentiable : $\dfrac{\partial p}{\partial \theta}(x;\theta)$ et $\dfrac{\partial^2 p}{\partial \theta^2}(x;\theta)$ existent pour tous $x \in \Delta$ et $\theta \in \Theta$.
3. On suppose que les fonctions $\dfrac{\partial p}{\partial \theta}$ et $\dfrac{\partial^2 p}{\partial \theta^2}$ sont intégrables pour tout $\theta \in \Theta$, et que les dérivées peuvent s'effectuer sous le signe somme. Ainsi pour tout $\theta \in \Theta$ et $A \subset \mathcal{X}$ borélien, on a :

$$\frac{\partial}{\partial \theta} \int_A p(x;\theta) dx = \int_A \frac{\partial}{\partial \theta} p(x;\theta) dx,$$
$$\text{et} \quad \frac{\partial^2}{\partial \theta^2} \int_A p(x;\theta) dx = \int_A \frac{\partial^2}{\partial \theta^2} p(x;\theta) dx.$$

Le vecteur du **score** de X_1, défini par $V_1 = \dfrac{\partial \log p}{\partial \theta}(X_1;\theta)$, joue un rôle particulier. S'il est intégrable il est centré. En effet, on a :

$$\begin{aligned}
\mathbb{E}_\theta\left[\frac{\partial \log p}{\partial \theta}(X_1;\theta)\right] &= \int_{\mathcal{X}} \frac{\partial \log p}{\partial \theta}(x;\theta) p(x;\theta)\, dx \\
&= \int_{\mathcal{X}} \frac{\partial p}{\partial \theta}(x;\theta)\, dx \\
&= \frac{\partial}{\partial \theta} \int_{\mathcal{X}} p(x;\theta)\, dx = 0.
\end{aligned}$$

Le vecteur du score de l'échantillon de taille n est défini par :

$$V_n = \frac{\partial \log p_n}{\partial \theta}(X_1, \ldots, X_n;\theta) = \frac{\partial}{\partial \theta} L_n(X_1, \ldots, X_n;\theta) = \sum_{i=1}^n \frac{\partial \log p}{\partial \theta}(X_i;\theta).$$

On rappelle que cette quantité intervient dans la recherche du maximum de vraisemblance. En effet, l'estimateur du maximum de vraisemblance maximise la log-vraisemblance $L_n(X;\theta)$. En particulier si le maximum est atteint à l'intérieur de

Θ, alors l'estimateur du maximum de vraisemblance annule la dérivée de $L_n(X;\theta)$ en θ. Donc l'estimateur du maximum de vraisemblance annule le vecteur du score.

Une notion importante liée au vecteur du score est sa variance (ou matrice de covariance dans le cas vectoriel) notée $I(\theta)$. Elle est définie dans le cas réel par :

$$\boxed{I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \log p}{\partial\theta}(X_1;\theta)\right)^2\right]} = \mathbb{E}_\theta\left[\left(\frac{\partial L_1}{\partial\theta}(X_1;\theta)\right)^2\right]$$

et dans le cas vectoriel par :

$$\boxed{I(\theta) = \mathbb{E}_\theta\left[\frac{\partial \log p}{\partial\theta}(X_1;\theta)\left(\frac{\partial \log p}{\partial\theta}(X_1;\theta)\right)^t\right].}$$

La quantité $I(\theta)$ est appelée l'**information de Fisher**. L'information de Fisher $I_n(\theta)$ de l'échantillon de taille n est la matrice de covariance du vecteur du score V_n. Par indépendance il vient :

$$\boxed{I_n(\theta) = nI(\theta).}$$

Enfin, on suppose :

4. L'information de Fisher existe i.e. le vecteur du score est de carré intégrable.

Proposition VIII.13. *Sous les hypothèses 1-4 de ce paragraphe, on a :*

$$\boxed{I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log p(X_1;\theta)}{\partial^2\theta}\right].}$$

Démonstration. On donne la démonstration dans le cas où le paramètre θ est réel. La démonstration du cas vectoriel est similaire. On remarque que :

$$\frac{\partial^2}{\partial\theta^2}\log p(X_1;\theta) = \frac{1}{p(X_1;\theta)}\frac{\partial^2}{\partial\theta^2}p(X_1;\theta) - \left(\frac{\partial}{\partial\theta}\log p(X_1;\theta)\right)^2.$$

En prenant l'espérance par rapport à $\mathbb{P}_\theta$, il vient :

$$\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p(X_1;\theta)\right] = \int_{\mathcal{X}}\frac{\partial^2}{\partial\theta^2}p(x;\theta)\,dx - I(\theta).$$

Grâce à l'hypothèse 3, on a $\displaystyle\int_{\mathcal{X}}\frac{\partial^2}{\partial\theta^2}p(x;\theta)\,dx = \frac{\partial^2}{\partial\theta^2}\int_{\mathcal{X}}p(x;\theta)\,dx = 0$. On en déduit la proposition. □

Définition VIII.14. *Si Θ est un ouvert et si les quatre hypothèses ci-dessus sont vérifiées, alors on dit que le modèle est* ***régulier****.*

Exemple VIII.15. Le modèle d'échantillonnage gaussien à variance σ_0 connue : $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\right\}$ est régulier. Le score V_n de l'échantillon de taille n est $V_n = \dfrac{n}{\sigma_0^2}\left(\bar{X}_n - \mu\right)$, où $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est la moyenne empirique. L'information de Fisher est :

$$I(\mu) = \mathbb{E}_\mu\left[\frac{1}{\sigma_0^4}(X_1-\mu)^2\right] = \frac{1}{\sigma_0^2} \quad \text{et} \quad I_n(\mu) = \frac{n}{\sigma_0^2}.$$

$\Diamond$

Exemple VIII.16. Le modèle d'échantillonnage gaussien à moyenne et variance inconnues $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\right\}$ est régulier. Le paramètre est $\theta = (\mu, \sigma)$. Pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on a :

$$\frac{\partial L_n}{\partial \mu}(x;\mu,\sigma) = \frac{n}{\sigma^2}\,\frac{1}{n}\sum_{i=1}^n (x_i - \mu) \quad \text{et} \quad \frac{\partial L_n}{\partial \sigma}(x;\mu,\sigma) = \frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma}.$$

Le vecteur de dimension 2 du score $V_n = (V_n(1), V_n(2))$ de l'échantillon de taille n est :

$$V_n(1) = \frac{n}{\sigma^2}(\bar{X}_n - \mu) \quad \text{et} \quad V_n(2) = \frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma}.$$

L'information de Fisher est :

$$I(\mu,\sigma) = \mathbb{E}_{\mu,\sigma}[V_1 V_1^t] = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \text{ et } I_n(\mu,\sigma) = \mathbb{E}_{\mu,\sigma}[V_n V_n^t] = \frac{n}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

$\Diamond$

Exemple VIII.17. Le modèle d'échantillonnage pour la loi de Bernoulli, $\mathcal{P} = \{\mathcal{B}(1,\theta); \theta \in]0,1[\}$, est régulier. La log-vraisemblance de l'échantillon est pour $x = (x_1, \ldots, x_n) \in \{0,1\}^n$:

$$L_n(x;\theta) = \log\left[\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}\right] = \left(\sum_{i=1}^n x_i\right)\log\frac{\theta}{1-\theta} + n\log(1-\theta).$$

On en déduit que le score de l'échantillon est :

$$V_n = \frac{n(\bar{X}_n - \theta)}{\theta(1-\theta)},$$

où $\bar{X}_n$ est la moyenne empirique. L'information de Fisher est :

$$I(\theta) = \mathbb{E}_\theta\left[V_1^2\right] = \frac{\mathrm{Var}_\theta(X_1)}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)} \quad \text{et} \quad I_n(\theta) = \frac{n}{\theta(1-\theta)}.$$

$\Diamond$

VIII.4.3 Borne FDCR

On considère un modèle **régulier**. Soit δ un estimateur de $g(\theta)$ de carré intégrable : $\mathbb{E}_\theta[|\delta|^2] < \infty$ pour tout $\theta \in \Theta$. On suppose de plus que, pour tout $\theta \in \Theta$, la fonction de $x \in \mathcal{X}^n$ définie par $\delta(x)\dfrac{\partial}{\partial\theta}p_n(x;\theta)$ est intégrable sur $\mathcal{X}^n$, et :

$$\frac{\partial}{\partial\theta}\int_{\mathcal{X}^n}\delta(x)p_n(x;\theta)\,dx = \int_{\mathcal{X}^n}\delta(x)\frac{\partial}{\partial\theta}p_n(x;\theta)\,dx.$$

On dit alors que δ est un estimateur **régulier** de $g(\theta)$. On peut alors minorer le risque quadratique de l'estimateur δ.

Théorème VIII.18 (Fréchet-Darmois-Cramer-Rao)**.** *On suppose que le paramètre θ est réel et que la fonction g est à valeurs réelles et de classe C^1. Soit δ un estimateur de carré intégrable, régulier et sans biais de $g(\theta)$. On suppose que $I(\theta) > 0$. Alors si la taille de l'échantillon est n, pour tout $\theta \in \Theta$, on a :*

$$\boxed{R(\delta,\theta) = \mathrm{Var}_\theta(\delta) \geq \frac{1}{n}\frac{g'(\theta)^2}{I(\theta)}.}$$

La borne $\dfrac{1}{n}\dfrac{g'(\theta)^2}{I(\theta)}$ est appelée la borne FDCR du modèle d'échantillonnage.

Dans le cas où g est l'identité, on obtient le corollaire suivant.

Corollaire VIII.19. *Si δ est un estimateur de carré intégrable, régulier et sans biais du paramètre réel θ construit à partir d'un échantillon de taille n, alors on a : pour tout $\theta \in \Theta \subset \mathbb{R}$,*

$$\boxed{R(\delta,\theta) = \mathbb{E}_\theta\left[(\delta-\theta)^2\right] \geq \frac{1}{nI(\theta)}\,.}$$

Démonstration du théorème VIII.18. Comme δ est un estimateur sans biais, on a :

$$g(\theta) = \mathbb{E}_\theta[\delta] = \int_{\mathcal{X}^n}\delta(x)p_n(x;\theta)\,dx.$$

En dérivant par rapport à θ, il vient :

$$\begin{aligned} g'(\theta) &= \int_{\mathcal{X}^n} \delta(x) \frac{\partial \log p_n(x;\theta)}{\partial \theta} p_n(x;\theta)\, dx \\ &= \mathbb{E}_\theta \left[\delta(X) \frac{\partial L_n}{\partial \theta}(X;\theta) \right] \\ &= \mathbb{E}_\theta \left[(\delta(X) - g(\theta)) \frac{\partial L_n}{\partial \theta}(X;\theta) \right]. \end{aligned}$$

car $\mathbb{E}_\theta \left[\frac{\partial L_n}{\partial \theta}(X;\theta) \right] = 0$. On en déduit en utilisant l'inégalité de Cauchy-Schwarz, que :

$$\left| g'(\theta) \right| \leq \mathbb{E}_\theta \left[\left| (\delta(X) - g(\theta)) \frac{\partial L_n}{\partial \theta}(X;\theta) \right| \right] \leq \sqrt{\mathrm{Var}_\theta(\delta(X))} \sqrt{I_n(\theta)}.$$

Enfin l'égalité $I_n(\theta) = nI(\theta)$ permet de conclure. □

On donne également la version vectorielle de ce théorème, dont la démonstration est similaire. On suppose que le paramètre θ est de dimension p : $\theta \in \Theta \subset \mathbb{R}^p$ et que la fonction $g : \Theta \to \mathbb{R}^k$ est de classe C^1. On note $\frac{\partial g}{\partial \theta}(\theta) = \left(\frac{\partial g_i}{\partial \theta_j}(\theta); 1 \leq i \leq k, 1 \leq j \leq p \right)$ qui est une matrice de taille $p \times k$.

Théorème VIII.20. *Soit δ un estimateur de carré intégrable, régulier et sans biais de $g(\theta)$. On suppose que g est de classe C^1 et que la matrice $I(\theta)$ est inversible. Alors si la taille de l'échantillon est n, pour tout $\theta \in \Theta$, on a :*

$$\boxed{R(\delta,\theta) = \mathrm{Var}_\theta(\delta) \geq \frac{1}{n} \frac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1} \left(\frac{\partial g}{\partial \theta}(\theta) \right)^t.}$$

Définition VIII.21. *Un estimateur sans biais atteignant la borne FDCR est dit* ***efficace****.*

Un estimateur efficace est donc préférable à tout autre estimateur sans biais. Toutefois il peut exister un estimateur biaisé qui lui soit préférable (cf. l'estimateur de la variance dans le paragraphe suivant).

La borne FDCR peut ne pas être atteinte. En fait on peut caractériser les modèles et les paramètres pour lesquels il existe un estimateur qui atteint la borne FDCR (cf. le paragraphe VIII.5.3).

VIII.4.4 Modèle gaussien

On distingue les trois cas : moyenne inconnue et variance connue ; moyenne connue et variance inconnue ; moyenne et variance inconnues.

Moyenne inconnue et variance connue.

On considère un modèle gaussien à variance connue et moyenne inconnue : $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$. La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur efficace de μ.

En effet cet estimateur est régulier et sans biais. De plus on a :

$$\mathrm{Var}_\mu(\bar{X}_n) = \mathbb{E}_\mu\left[\left(\bar{X}_n - \mu\right)^2\right] = \frac{\mathrm{Var}_\mu X_1}{n} = \frac{\sigma_0^2}{n}.$$

La borne FDCR du modèle est $(nI(\mu))^{-1}$, où $I(\mu) = 1/\sigma_0^2$ (voir l'exemple VIII.15). La borne FDCR σ_0^2/n est donc atteinte. La moyenne empirique est donc un estimateur efficace de μ. On peut en fait montrer qu'il n'existe pas d'estimateur de carré intégrable préférable à $\bar{X}_n$ dans ce cas très précis (cf. l'exercice VIII.11).

Moyenne connue et variance inconnue.

Le modèle gaussien à moyenne connue $\mathcal{P} = \{\mathcal{N}(\mu_0, \sigma^2); \sigma > 0\}$ est un modèle régulier. On recherche un estimateur efficace de la variance $g(\sigma) = \sigma^2$. Le vecteur du score de l'échantillon de taille n est :

$$\frac{\partial}{\partial\sigma} L_n(X;\sigma) = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^3} - \frac{n}{\sigma},$$

l'information de Fisher est :

$$I(\sigma) = -\mathbb{E}_\sigma\left[\frac{\partial^2 L_1(X_1;\sigma)}{\partial\sigma^2}\right] = \mathbb{E}_\sigma\left[3\frac{(X_1-\mu_0)^2}{\sigma^4} - \frac{1}{\sigma^2}\right] = \frac{2}{\sigma^2}.$$

La variance empirique $\hat{\sigma}_n^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^n (X_i - \mu_0)^2$ est un estimateur régulier de σ^2. Il est sans biais et on calcule :

$$\begin{aligned}\mathrm{Var}_\sigma(\hat{\sigma}_n^2) &= \mathbb{E}_\sigma\left[\left(\frac{1}{n}\sum_{i=1}^n (X_i-\mu_0)^2 - \sigma^2\right)^2\right] \\ &= \frac{1}{n}\,\mathrm{Var}_\sigma((X_1-\mu_0)^2) = \frac{2\sigma^4}{n}.\end{aligned}$$

Comme $g(\sigma) = \sigma^2$, on a $g'(\sigma) = 2\sigma$. La borne FDCR est donc :

$$\frac{1}{n}\, g'(\sigma)^2\, \frac{1}{I(\sigma)} = \frac{2\sigma^4}{n}.$$

L'estimateur $\hat{\sigma}_n^2$ est donc un estimateur efficace de σ^2. Toutefois, parmi les estimateurs de la forme $c\hat{\sigma}_n^2$, un calcul élémentaire montre que le risque quadratique est minimal pour $c = n/(n+2)$. L'estimateur (biaisé) $\dfrac{n}{n+2}\hat{\sigma}_n^2$ est donc préférable à $\hat{\sigma}_n^2$, bien que ce dernier soit efficace.

Moyenne inconnue et variance inconnue.

Le modèle d'échantillonnage gaussien $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\right\}$ est un modèle régulier. Le paramètre est $\theta = (\mu, \sigma)$. La moyenne empirique $\hat{\mu}_n = \bar{X}_n$ est un estimateur sans biais de μ. On remarque que $\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ n'est pas un estimateur car il dépend de μ, et donc du paramètre θ. On vérifie que l'estimateur du maximum de vraisemblance de (μ, σ^2) est $(\hat{\mu}_n, \tilde{\sigma}_n^2)$, où :

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2.$$

Ce dernier est biaisé car :

$$\mathbb{E}_{\mu,\sigma}\left[\tilde{\sigma}_n^2\right] = \sigma^2\left(1 - \frac{1}{n}\right) = \frac{n-1}{n}\sigma^2.$$

On considère donc l'estimateur sans biais :

$$V_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

L'estimateur $(\hat{\mu}_n, V_n)$ est un estimateur régulier et sans biais de $g(\mu, \sigma) = (\mu, \sigma^2)$. On sait (cf. chapitre VI.2) que les variables aléatoires $\hat{\mu}_n$ et $(n-1)\dfrac{V_n}{\sigma^2}$ sont indépendantes et de lois respectives $\mathcal{N}(\mu, \sigma^2/n)$ et $\chi^2(n-1)$. La matrice de covariance du couple $(\hat{\mu}_n, V_n)$ est donc :

$$K = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{pmatrix}.$$

On a déjà calculé l'information de Fisher $I(\theta) = I(\mu, \sigma) = \dfrac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ (cf. exemple VIII.16). La borne FDCR vectorielle est :

$$J = \frac{1}{n} \frac{\partial g}{\partial \theta}(\mu, \sigma) I(\mu, \sigma)^{-1} \left(\frac{\partial g}{\partial \theta}(\mu, \sigma) \right)^t .$$

On obtient donc :

$$J = \begin{pmatrix} 1 & 0 \\ 0 & 2\sigma \end{pmatrix} \frac{\sigma^2}{n} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2\sigma \end{pmatrix} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}.$$

Le couple $(\hat{\mu}_n, V_n)$ n'est donc pas un estimateur efficace, car la borne FDCR n'est pas atteinte : $K \neq J$. En revanche, on verra au prochain paragraphe qu'il est toutefois préférable à tout autre estimateur sans biais. Il n'existe donc pas d'estimateur efficace de (μ, σ^2).

VIII.5 Amélioration d'estimateurs

Le but de ce paragraphe est de résumer les données et d'améliorer au sens du risque quadratique les estimateurs.

On considère un modèle paramétrique d'échantillonnage de taille n : les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et de loi $\mathrm{P}_\theta \in \mathcal{P}$. On ne suppose pas dans ce paragraphe que le modèle est régulier.

VIII.5.1 Statistiques exhaustives, statistiques totales

Une statistique permet de résumer les données si elle contient toute l'information sur le paramètre. La définition suivante donne un sens mathématique à cette intuition.

Définition VIII.22. *Une statistique S est **exhaustive** si la loi conditionnelle de l'échantillon $(X_1, \ldots, X_n)$ sachant S est indépendante du paramètre θ.*

Exemple. Soit le modèle d'échantillonnage de Bernoulli $\mathcal{P} = \{\mathcal{B}(1, \theta); \theta \in]0, 1[\}$. On montre que la statistique $S = \sum_{i=1}^n X_i$ est exhaustive. Pour $(k_1, \ldots, k_n) \in \{0, 1\}^n$, on a :

$$\mathbb{P}_\theta \left(X_1 = k_1, \ldots, X_n = k_n | S = k \right) = \frac{\prod_{i;k_i=1} \theta \prod_{j;k_j=0} (1-\theta)}{C_n^k \theta^k (1-\theta)^{n-k}} = \frac{1}{C_n^k}$$

si $\sum_{i=1}^n k_i \neq k$ et $\mathbb{P}_\theta \left(X_1 = k_1, \ldots, X_n = k_n | S = k \right) = 0$ sinon. On en déduit que la loi conditionnelle de $(X_1, \ldots, X_n)$ sachant S est la loi uniforme sur $\{(k_1, \ldots, k_n) \in \{0, 1\}^n; \sum_{i=1}^n k_i = S\}$. La statistique S est donc exhaustive. ◊

Exercice VIII.2.
On considère le modèle d'échantillonnage $\mathcal{P} = \{\Gamma(\lambda, \alpha_0); \lambda > 0\}$, le paramètre α_0 étant connu. Calculer la loi de $S = \sum_{i=1}^n X_i$. Calculer la loi de $\left(S, \frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}\right)$. Montrer que la statistique S est exhaustive. La loi de $\left(\frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}\right)$ porte le nom de loi de Dirichlet. △

Le théorème suivant dont la démonstration est admise permet d'exhiber des statistiques exhaustives.

Théorème VIII.23 (de factorisation)**.** *La statistique $S = S(X_1, \ldots, X_n)$ est exhaustive si et seulement si la densité $p_n(x_1, \ldots, x_n; \theta)$ de la loi de l'échantillon $(X_1, \ldots, X_n)$ se factorise de la façon suivante : il existe des fonctions Ψ et ℓ telles que : pour tout $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $\theta \in \Theta$,*

$$\boxed{p_n(x; \theta) = \Psi(S(x), \theta)\ell(x).}$$

Exemple. On reprend l'exercice VIII.2. La densité de l'échantillon est :

$$p_n(x_1, \ldots, x_n; \lambda) = \frac{\lambda^{n\alpha_0}}{\Gamma(\alpha_0)^n} \prod_{i=1}^n \left(x_i^{\alpha_0 - 1} \mathbf{1}_{\{x_i > 0\}}\right) \mathrm{e}^{-\lambda \sum_{i=1}^n x_i}.$$

Elle se factorise avec $S(x) = \sum_{i=1}^n x_i$, $\ell(x) = \prod_{i=1}^n \left(x_i^{\alpha_0 - 1} \mathbf{1}_{\{x_i > 0\}}\right)$ et $\Psi(s, \lambda) = \frac{\lambda^{n\alpha_0}}{\Gamma(\alpha_0)^n} \mathrm{e}^{-\lambda s}$. On en déduit que la statistique $S = \sum_{i=1}^n X_i$ est exhaustive. Cette méthode évite le calcul de la loi conditionnelle. ◊

Exercice VIII.3.
Démontrer le théorème dans le cas de lois discrètes. △

Exemple VIII.24. On considère le modèle uniforme $\mathcal{P} = \left\{U_{[0,\theta]}; \theta > 0\right\}$. La densité de l'échantillon de taille n est :

$$p_n(x_1, \ldots, x_n; \theta) = \frac{1}{\theta^n} \mathbf{1}_{\{\max_{1 \le i \le n} x_i \le \theta\}} \mathbf{1}_{\{\min_{1 \le i \le n} x_i \ge 0\}}.$$

On en déduit que la statistique $\max_{1 \le i \le n} X_i$ est exhaustive. ◊

La notion de statistique totale introduite ci-dessous sera utilisée dans le paragraphe VIII.5.2 pour améliorer les estimateurs.

Définition VIII.25. *Une statistique S est **totale** si $\mathbb{E}_\theta[|h(S)|] < \infty$ et $\mathbb{E}_\theta[h(S)] = 0$ pour tout $\theta \in \Theta$ implique $h(S) = 0$ $\mathbb{P}_\theta$-p.s. pour tout $\theta \in \Theta$.*

Exemple VIII.26. On considère le modèle de Bernoulli $\mathcal{P} = \{\mathcal{B}(1,\theta); \theta \in]0,1[\}$. On montre que la statistique $S = \sum_{i=1}^{n} X_i$ est totale. Soit h telle que : pour tout $\theta \in]0,1[$,

$$0 = \mathbb{E}_\theta[h(S)] = (1-\theta)^n \sum_{k=0}^{n} C_n^k h(k) \left(\frac{\theta}{1-\theta}\right)^k.$$

La fonction $x \mapsto P(x) = \sum_{k=0}^{n} C_n^k h(k) x^k$ est un polynôme de degré n, nul pour $x \in]0,\infty[$. Les coefficients de ce polynôme sont donc nuls. Cela implique que $h(k) = 0$ pour tout $k \in \{0,\dots,n\}$. On en déduit que $h(S) = 0$ $\mathbb{P}_\theta$-p.s. pour tout $\theta \in]0,1[$. La statistique S est donc totale.

En revanche la statistique $S = (X_1,\dots,X_n)$ n'est pas totale. En effet la fonction h définie par $(x_1,\dots,x_n) \mapsto x_1 - x_2$ vérifie $\mathbb{E}_\theta[h(S)] = 0$ pour tout $\theta \in]0,1[$, alors que $\mathbb{P}_\theta(X_1 - X_2 > 0) = \theta(1-\theta) > 0$. ♢

Remarque VIII.27. On dit qu'une statistique exhaustive S est **minimale** si pour toute statistique exhaustive T, il existe une fonction mesurable (déterministe) h telle que $S = h(T)$. En un certain sens, la statistique exhaustive minimale est le plus petit résumé des données qui contienne toute l'information sur le paramètre. On peut montrer qu'une statistique exhaustive totale est alors minimale. ♢

VIII.5.2 Estimateurs améliorés de Rao-Blackwell

Le théorème de Rao et Blackwell permet d'améliorer un estimateur en utilisant une statistique exhaustive.

On considère l'échantillon $X_1,\dots,X_n$. La loi de $(X_1,\dots,X_n)$ sachant une statistique exhaustive S est indépendante de θ. En particulier l'espérance conditionnelle $\mathbb{E}_\theta[\delta|S] = \mathbb{E}_\theta[\delta(X_1,\dots,X_n)|S]$ ne dépend plus de θ. Pour cette raison, on supprime l'indice θ et on la note $\mathbb{E}[\delta|S]$.

Théorème VIII.28 (Rao-Blackwell)**.** *Soit δ un estimateur de carré intégrable de $g(\theta)$. Si S est une statistique exhaustive, alors l'estimateur $\delta_S = \mathbb{E}[\delta|S]$ est un estimateur préférable à δ. Si de plus δ est sans biais, alors δ_S est aussi sans biais.*

Démonstration. L'estimateur δ_S est de carré intégrable car δ est de carré intégrable (voir le lemme II.50 vrai en général). Ils ont même biais car :

$$\mathbb{E}_\theta[\delta_S] = \mathbb{E}_\theta\big[\mathbb{E}[\delta|S]\big] = \mathbb{E}_\theta[\delta].$$

Il vient :

$$\begin{aligned} R(\delta,\theta) - R(\delta_S,\theta) &= \mathbb{E}_\theta\big[(\delta - g(\theta))(\delta - g(\theta))^t\big] - \mathbb{E}_\theta\big[(\delta_S - g(\theta))(\delta_S - g(\theta))^t\big] \\ &= \mathbb{E}_\theta\big[\delta\delta^t - \delta_S{\delta_S}^t\big] \\ &= \mathbb{E}_\theta[(\delta - \delta_S)(\delta - \delta_S)^t], \end{aligned}$$

où pour la dernière égalité on utilise les propriétés de l'espérance conditionnelle qui assurent que $\mathbb{E}_\theta[\delta{\delta_S}^t] = \mathbb{E}_\theta[\mathbb{E}[\delta|S]{\delta_S}^t] = \mathbb{E}[\delta_S{\delta_S}^t]$.

Comme la matrice symétrique $(\delta - \delta_S)(\delta - \delta_S)^t$ est positive, on en déduit que $R(\delta,\theta) \geq R(\delta_S,\theta)$. L'estimateur δ_S est donc préférable à l'estimateur δ. □

Définition VIII.29. *Soit δ un estimateur de $g(\theta)$ de carré intégrable et sans biais. On dit que δ est un estimateur **optimal** dans la famille des estimateurs sans biais de $g(\theta)$ s'il est préférable à tout autre estimateur sans biais de $g(\theta)$.*

Remarque. Un estimateur efficace sans biais est un estimateur optimal dans la famille des estimateurs sans biais. Mais tous les estimateurs optimaux ne sont pas des estimateurs efficaces. ♢

Le théorème suivant donne une méthode pour construire des estimateurs optimaux.

Théorème VIII.30 (Lehman-Sheffé)**.** *Soit δ un estimateur sans biais de $g(\theta)$ et de carré intégrable. Si S est une statistique exhaustive et totale, alors l'estimateur amélioré δ_S de Rao-Blackwell de δ est optimal dans la famille des estimateurs sans biais de $g(\theta)$.*

Démonstration. Soit δ^* un autre estimateur sans biais de $g(\theta)$ et δ_S^* son amélioré de Rao-Blackwell. On a donc $\mathbb{E}_\theta[\delta_S - \delta_S^*] = 0$ pour tout $\theta \in \Theta$. On remarque que δ_S et δ_S^* sont des fonctions de S. Comme la statistique S est totale, on a $\delta_S = \delta_S^*$ $\mathbb{P}_\theta$-p.s. pour tout $\theta \in \Theta$. On en déduit donc que :

$$R(\delta^*,\theta) \geq R(\delta_S^*,\theta) = R(\delta_S,\theta).$$

L'estimateur δ_S est donc optimal. □

Exemple. On considère le modèle uniforme $\mathcal{P} = \big\{U_{[0,\theta]}; \theta > 0\big\}$. On a vu à l'exercice VIII.24 que la statistique $\hat{\theta}_n = \max_{1\leq i\leq n} X_i$ est exhaustive. On montre qu'elle est totale. On connaît la densité de la loi de $\hat{\theta}_n$ (cf. exemple VIII.12). Soit h intégrable telle que $\mathbb{E}_\theta[h(\hat{\theta}_n)] = 0$ pour tout $\theta > 0$. Cela implique que pour tout $\theta > 0$:

$$0 = n\theta^{-n} \int_0^\theta h(t) t^{n-1}\, dt = n\theta^{-n} F(\theta) \quad \text{où} \quad F(\theta) = \int_0^\theta h(t) t^{n-1} dt.$$

Si la fonction h est continue, alors F est de classe C^1 et par différenciation de F, on a $h = 0$ sur $\mathbb{R}^+$. Si h n'est pas continue, on admet que $F = 0$ implique $h = 0$ sur $\mathbb{R}^+$ sauf sur un ensemble de mesure de Lebesgue nul. Cela signifie que $h = 0$ $\mathbb{P}_\theta$-p.s. pour tout $\theta > 0$. La statistique $\hat{\theta}_n$ est donc totale. L'estimateur $\frac{n+1}{n}\hat{\theta}_n$ est sans biais. Grâce au théorème de Lehman-Sheffé, il s'agit d'un estimateur optimal dans la famille des estimateurs sans biais. L'exercice VIII.1 montre toutefois que cet estimateur est inadmissible. On remarque également que le modèle n'étant pas régulier, il n'y a pas de borne FDCR. ◊

VIII.5.3 Le modèle exponentiel

Les modèles exponentiels sont des modèles paramétriques assez généraux pour lesquels ont sait exhiber des estimateurs efficaces.

Définition VIII.31. *Une famille paramétrique de probabilités $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ forme un modèle **exponentiel** s'il existe des fonctions C, h, $Q = (Q^{(1)}, \ldots, Q^{(r)})$ et $S = (S^{(1)}, \ldots, S^{(r)})$ telles que la densité est de la forme :*

$$\boxed{p(x_1; \theta) = C(\theta) h(x_1)\, \mathrm{e}^{\sum_{j=1}^r Q^{(j)}(\theta) S^{(j)}(x_1)}}.$$

*Si les fonctions $Q^{(1)}, \ldots, Q^{(r)}$ sont linéairement indépendantes, alors la statistique $S = (S^{(1)}, \ldots, S^{(r)})$ est dite **canonique**.*

La statistique $S_n = (S_n^{(1)}, \ldots, S_n^{(r)})$ avec $S_n^{(j)} = \sum_{i=1}^n S^{(j)}(X_i)$ est la statistique canonique de l'échantillon de taille n. La densité de l'échantillon de taille n se met sous la forme :

$$p_n(x; \theta) = \prod_{i=1}^n p(x_i; \theta) = C(\theta)^n \left(\prod_{i=1}^n h(x_i) \right) \mathrm{e}^{\sum_{j=1}^r Q^{(j)}(\theta) S_n^{(j)}(x_1, \ldots, x_n)},$$

où $x = (x_1, \ldots, x_n)$ et $S_n^{(j)}(x_1, \ldots, x_n) = \sum_{i=1}^n S^{(j)}(x_i)$.

Exemple. Les familles de lois de Bernoulli $\mathcal{P} = \{\mathcal{B}(1, \theta); \theta \in]0, 1[\}$, de lois exponentielles $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$, de lois gamma $\mathcal{P} = \{\Gamma(\lambda, \alpha); \lambda > 0,\ \alpha > 0\}$ et de lois gaussiennes $\mathcal{P} = \{\mathcal{N}(\mu, \sigma); \mu \in \mathbb{R}, \sigma > 0\}$ sont par exemple des familles exponentielles. ◊

Grâce au théorème de factorisation VIII.23, on en déduit la proposition suivante.

Proposition VIII.32. *La statistique canonique du modèle d'échantillonnage est exhaustive.*

On peut réécrire la densité de P_θ sous sa forme dite **naturelle**. Pour cela on pose $\lambda = Q(\theta)$ soit $\lambda_j = Q^{(j)}(\theta)$ pour $1 \leq j \leq n$. On a alors avec $\varphi(\lambda) = -\log(C(\theta))$:

$$\boxed{p(x_1;\lambda) = h(x_1)\,\mathrm{e}^{\sum_{j=1}^{r}\lambda_j S^{(j)}(x_1)-\varphi(\lambda)}\,.}$$

On admet le théorème suivant.

Théorème VIII.33. *On reprend les notations de la définition VIII.31. Si $\Lambda = \{\lambda = Q(\theta), \theta \in \Theta\}$ est un ouvert de $\mathbb{R}^r$, alors la statistique $S_n = \sum_{i=1}^n S(X_i)$ de l'échantillon de taille n est totale.*

D'après la remarque VIII.27, la statistique S_n est minimale c'est-à-dire elle est fonction de toute autre statistique exhaustive.

Exemple VIII.34. Le modèle d'échantillonnage $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$, est exponentiel. La densité de l'échantillon de taille n est :

$$p_n(x;\lambda) = \lambda^n\,\mathrm{e}^{-\lambda\sum_{i=1}^n x_i}\prod_{i=1}^n \mathbf{1}_{\{x_i>0\}},$$

où $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$. Le modèle est en fait sous sa forme naturelle. La statistique $S_n = \sum_{i=1}^n X_i$ est exhaustive minimale et totale. ◊

Exemple. Soit le modèle d'échantillonnage de loi Γ : $\mathcal{P} = \{\Gamma(\lambda,\alpha); \lambda > 0, \alpha > 0\}$, où :

$$p_1(x_1;\lambda,\alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x_1^{\alpha-1}\,\mathrm{e}^{-\lambda x_1}\,\mathbf{1}_{\{x_1>0\}}.$$

Il s'agit d'un modèle exponentiel. La densité de l'échantillon de taille n est :

$$p_n(x;\lambda,\alpha) = \frac{\lambda^n}{\Gamma(\alpha)^n}\,\mathrm{e}^{-\lambda\sum_{i=1}^n x_i+(\alpha-1)\log(\prod_{i=1}^n x_i)}\prod_{i=1}^n \mathbf{1}_{\{x_i>0\}}.$$

La statistique $S_n = (\sum_{i=1}^n X_i, \sum_{i=1}^n \log X_i)$ est exhaustive minimale et totale. ◊

On suppose que φ est de classe C^2. Dans ce cas le modèle est régulier. On calcule pour le modèle sous sa forme naturelle, le vecteur du score et l'information de Fisher. Le vecteur du score de X_1 est :

$$V_1 = \frac{\partial \log p(X_1;\lambda)}{\partial \lambda} = \left(S^{(1)}(X_1) - \frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \dots, S^{(r)}(X_1) - \frac{\partial \varphi(\lambda)}{\partial \lambda_r} \right).$$

L'information de Fisher est :

$$I(\lambda) = -\mathbb{E}_\lambda \left[\frac{\partial^2 \log p(X_1;\lambda)}{\partial \lambda^2} \right] = \frac{\partial^2 \varphi(\lambda)}{\partial \lambda^2} = \mathrm{Jac}[\varphi](\lambda).$$

L'information de Fisher est ainsi le Jacobien de la fonction φ. Ce résultat est vrai pour la forme naturelle. On remarque que l'on ne calcule pas $I(\theta)$. Le vecteur du score de l'échantillon de taille n est donc :

$$\begin{aligned} V_n &= \frac{\partial \log p_n(X_1,\dots,X_n;\lambda)}{\partial \lambda} \\ &= \left(\sum_{i=1}^n S^{(1)}(X_i) - n\frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \dots, \sum_{i=1}^n S^{(r)}(X_i) - \frac{\partial \varphi(\lambda)}{\partial \lambda_r} \right) \\ &= \left(S_n^{(1)} - n\frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \dots, S_n^{(r)} - \frac{\partial \varphi(\lambda)}{\partial \lambda_r} \right). \end{aligned}$$

Comme le vecteur du score est centré, on en déduit que $S_n/n = \left(S_n^{(1)}, \dots, S_n^{(r)}\right)/n$ est un estimateur sans biais de :

$$\frac{\partial \varphi(\lambda)}{\partial \lambda} = \left(\frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \dots, \frac{\partial \varphi(\lambda)}{\partial \lambda_r} \right).$$

En fait il s'agit de l'estimateur du maximum de vraisemblance de $\dfrac{\partial \varphi(\lambda)}{\partial \lambda}$.

Proposition VIII.35. *L'estimateur S_n/n est un estimateur efficace de $\partial\varphi(\lambda)/\partial\lambda$.*

Démonstration. On a :

$$\begin{aligned} \mathrm{Var}_\lambda(S_n/n) &= \mathbb{E}\left[\left(\frac{S_n}{n} - \frac{\partial \varphi}{\partial \lambda}(\lambda) \right) \left(\frac{S_n}{n} - \frac{\partial \varphi}{\partial \lambda}(\lambda) \right)^t \right] \\ &= \frac{1}{n^2}\, \mathbb{E}\left[\frac{\partial \log p_n(X_1,\dots,X_n;\lambda)}{\partial \lambda} \left(\frac{\partial \log p_n(X_1,\dots,X_n;\lambda)}{\partial \lambda} \right)^t \right] \\ &= \frac{1}{n^2}\, I_n(\lambda) \\ &= \frac{1}{n}\, \frac{\partial^2 \varphi(\lambda)}{\partial \lambda^2}. \end{aligned}$$

La borne FDCR de l'estimation de $g(\lambda) = \dfrac{\partial\varphi(\lambda)}{\partial\lambda}$ est :

$$\frac{1}{n}\frac{\partial g(\lambda)}{\partial\lambda} I^{-1}(\lambda) \left(\frac{\partial g(\lambda)}{\partial\lambda}\right)^t = \frac{1}{n}\frac{\partial^2\varphi(\lambda)}{\partial\lambda^2}\left(\frac{\partial^2\varphi(\lambda)}{\partial\lambda^2}\right)^{-1}\frac{\partial^2\varphi(\lambda)}{\partial\lambda^2} = \frac{1}{n}\, I(\lambda).$$

□

Remarque. On peut montrer que s'il existe un estimateur W_n, de l'échantillon de taille n, non biaisé de $g(\theta)$, où g est un C^1 difféomorphisme, qui atteint la borne FDCR pour tout $\theta \in \Theta$ ouvert, alors si le modèle est régulier et si W_n est régulier, la famille $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ est exponentielle. De plus W_n est une fonction affine de S_n/n. ◊

Exemple VIII.36. On considère un modèle gaussien à moyenne connue et variance inconnue : $\mathcal{P} = \{\mathcal{N}(\mu_0, \sigma); \sigma > 0\}$. On recherche un estimateur sans biais de σ. La densité de l'échantillon de taille n est :

$$p_n(x;\sigma) = (2\pi\sigma^2)^{n/2}\, \mathrm{e}^{-\sum_{i=1}^n (x_i-\mu_0)^2/2\sigma^2}.$$

Il s'agit d'un modèle exponentiel. La statistique $T_n = \sum_{i=1}^n (X_i - \mu_0)^2$ est exhaustive minimale et totale. On cherche donc un estimateur sans biais de σ de la forme $h_n(T_n)$: pour tout $\sigma > 0$, $\mathbb{E}_\sigma[h_n(T_n)] = \sigma$. La fonction h_n est unique car la statistique T_n est totale. Comme T_n est un estimateur de σ^2, il est naturel de rechercher un estimateur de la forme $c_n\sqrt{T_n}$. La loi de T_n/σ^2 est un χ^2 à n degré de libertés. On en déduit que :

$$\mathbb{E}_\sigma[\sqrt{T_n}] = \sigma \int_0^\infty \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{\frac{n}{2}-1+\frac{1}{2}}\, \mathrm{e}^{-x/2}\; dx = \sigma\frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

L'estimateur $\tilde{\sigma}_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}\sqrt{\sum_{i=1}^n (X_i-\mu_0)^2}$ est donc sans biais. D'après le théorème de Lehman-Sheffé, il est optimal. L'estimateur T_n/n est un estimateur efficace (de σ^2, cf. le paragraphe VIII.4.4). D'après la remarque précédente, l'estimateur $\tilde{\sigma}_n$ ne peut donc pas être efficace. On le vérifie explicitement sur cet exemple. On connaît l'information de Fisher $I(\sigma) = 2/\sigma^2$ (cf. le paragraphe VIII.4.4). La borne FDCR est donc $\dfrac{1}{nI(\sigma)} = \dfrac{\sigma^2}{2n}$. Le calcul de la variance de l'estimateur $\tilde{\sigma}_n$ donne :

$$\mathrm{Var}_\sigma(\tilde{\sigma}_n) = \sigma^2\left[\frac{n}{2}\frac{\Gamma\left(\frac{n}{2}\right)^2}{\Gamma\left(\frac{n+1}{2}\right)^2} - 1\right].$$

En utilisant le développement limité de la fonction Γ :

$$\Gamma(z) = \mathrm{e}^{-z}\, z^{z-\frac{1}{2}} (2\pi)^{1/2} \left[1 + \frac{1}{12z} + \frac{1}{288z^2} + O\left(\frac{1}{z^3}\right)\right],$$

on obtient que $\mathrm{Var}_\sigma(\tilde{\sigma}_n) = \dfrac{\sigma^2}{2n} + \dfrac{\sigma^2}{8n^2} + O\left(\dfrac{1}{n^3}\right) > \dfrac{\sigma^2}{2n}$. En revanche on remarque que $\mathrm{Var}_\sigma(\tilde{\sigma}_n)$ et $\sigma^2/2n$ sont asymptotiquement équivalents (quand $n \to \infty$). ◊

VIII.6 Analyse asymptotique

Cette approche est valable lorsque la taille de l'échantillon est grande.

Exemple VIII.37. On considère un modèle de Bernoulli $\mathcal{P} = \{\mathcal{B}(1,\theta);\, \theta \in]0,1[\}$. Il s'agit d'un modèle exponentiel :

$$p_n(x;\theta) = \mathrm{e}^{\sum_{i=1}^n x_i \log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)}, \quad \text{où} \quad x = (x_1, \ldots, x_n) \in \{0,1\}.$$

La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est une statistique exhaustive totale minimale. De plus on a $\mathbb{E}_\theta[\bar{X}_n] = \theta$. Il s'agit donc d'un estimateur sans biais optimal de θ. On remarque que $\mathrm{Var}_\theta(\bar{X}_n) = \frac{1}{n}\,\theta(1-\theta)$. L'information de Fisher est :

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \log p_1(X_1;\theta)}{\partial\theta}\right)^2\right] = \frac{1}{\theta(1-\theta)}\,.$$

On en déduit que l'estimateur de la moyenne empirique est efficace. La loi forte des grands nombres assure qu'il est convergent. Le théorème central limite précise la vitesse de convergence :

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n\to\infty]{\textbf{en loi}} \mathcal{N}(0, \theta(1-\theta)).$$

On dit que $\theta(1-\theta)$ est la variance asymptotique. ◊

De nombreux estimateurs ont un comportement asymptotique similaire à celui de l'exercice précédent.

Définition VIII.38. *Une suite d'estimateurs $(\delta_n, n \in \mathbb{N}^*)$ de $g(\theta)$, où δ_n est une fonction de l'échantillon n, est* ***asymptotiquement normale*** *si pour tout $\mathrm{P}_{\theta_0} \in \mathcal{P}$, la suite $(\sqrt{n}\,(\delta_n - g(\theta_0)), n \in \mathbb{N}^*)$ converge en loi vers une loi gaussienne $\mathcal{N}(0, \Sigma(\theta_0))$. La matrice $\Sigma(\theta_0)$ est appelée* ***matrice de covariance asymptotique*** *(****variance asymptotique*** *si la fonction g est réelle) de la suite d'estimateurs.*

Par brièveté, on confondra estimateur et suite d'estimateurs.

Remarque VIII.39. Un estimateur δ_n de $g(\theta)$ asymptotiquement normal est faiblement convergent. En effet on a $\lim_{n\to\infty} n^{-1/2} = 0$. On déduit du théorème de Slutsky que la suite $\left(n^{-1/2}(\sqrt{n}(\delta_n - g(\theta))), n \in \mathbb{N}^*\right)$ converge en loi vers la constante 0. Cela implique donc la convergence en probabilité de la suite $\left(\delta_n - g(\theta), n \in \mathbb{N}^*\right)$ vers 0. ◊

VIII.6.1 Estimateurs de substitution

La méthode de substitution permet de construire aisément des estimateurs asymptotiquement normaux.

Proposition VIII.40. *Soit $\hat{\theta}_n$ un estimateur asymptotiquement normal de θ de variance asymptotique, ou de matrice de covariance asymptotique, $\Sigma(\theta)$. Soit g une fonction réelle ou vectorielle de classe C^1. Alors l'estimateur $g(\hat{\theta}_n)$ est un estimateur asymptotiquement normal de $g(\theta)$. Dans le cas réel (θ réel et g à valeurs dans $\mathbb{R}$), la variance asymptotique est :*

$$\boxed{g'(\theta)^2 \Sigma(\theta);}$$

dans le cas vectoriel la matrice de covariance asymptotique est :

$$\frac{\partial g}{\partial \theta}(\theta) \Sigma(\theta) \left(\frac{\partial g}{\partial \theta}(\theta)\right)^t.$$

Démonstration. Comme $\hat{\theta}_n$ est un estimateur asymptotiquement normal de θ, il est faiblement convergent d'après la remarque VIII.39. On démontre alors la proposition avec des arguments similaires à ceux utilisés dans la démonstration du corollaire VI.17. □

VIII.6.2 Estimateurs des moments

La proposition suivante découle directement du théorème central limite.

Proposition VIII.41. *Soit φ une fonction telle que, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[|\varphi(X_1)|^2] < \infty$ et $\mathbb{E}_\theta[\varphi(X_1)] = \theta$. Alors l'estimateur $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \varphi(X_i)$ de θ est convergent. Il est asymptotiquement normal et sa matrice de covariance asymptotique est la matrice de covariance de $\varphi(X_1)$.*

Enfin la méthode des moments décrite au chapitre VIII.3.2 fournit des estimateurs asymptotiquement normaux, pourvu que m^{-1} soit de classe C^1. Cela découle des propositions VIII.40 et VIII.41.

VIII.6.3 Estimateurs du maximum de vraisemblance

On admet le résultat important suivant.

Théorème VIII.42. *On suppose que le modèle est régulier (cf. définition VIII.14) et identifiable (cf. paragraphe VIII.1). L'estimateur du maximum de vraisemblance est asymptotiquement normal. De plus sa matrice de covariance asymptotique est l'inverse de l'information de Fisher. On a :*

$$\boxed{\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} \mathcal{N}(0, I(\theta)^{-1}), \quad \textit{sous } \mathbb{P}_\theta, \textit{ pour tout } \theta \in \Theta.}$$

On retrouve ce résultat dans l'exemple suivant.

Exemple. On considère le modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$. L'estimateur du maximum de vraisemblance de λ est :

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}\,.$$

La loi forte des grands nombres assure que l'estimateur est convergent. Le théorème central limite assure que $\sqrt{n}\left(\bar{X}_n - \dfrac{1}{\lambda}\right)$ converge en loi vers $\mathcal{N}(0, 1/\lambda^2)$. La méthode de substitution assure que $\sqrt{n}\left(\hat{\lambda}_n - \lambda\right)$ converge en loi vers $\mathcal{N}(0, \lambda^2)$ (prendre $g(x) = x^{-1}$, et on a $g'(\lambda^{-1})^2\,\frac{1}{\lambda^2} = \lambda^2$). Ainsi l'estimateur $\hat{\lambda}_n$ est asymptotiquement normal de variance asymptotique λ^2. Enfin l'information de Fisher de la loi exponentielle est :

$$I(\lambda) = \mathbb{E}_\lambda\left[-\frac{\partial^2 \log p_1(X_1;\lambda)}{\partial\lambda^2}\right] = \frac{1}{\lambda^2}\,.$$

On retrouve que l'estimateur du maximum de vraisemblance est asymptotiquement normal de variance asymptotique l'inverse de l'information de Fisher. ◊

Exercice VIII.4.
On reprend les notations de l'exemple ci-dessus. Montrer que $\hat{\lambda}_n$ est un estimateur biaisé de λ. Construire un estimateur optimal sans biais de λ. Vérifier, qu'il

n'est pas efficace. Montrer que la différence entre l'estimateur optimal et $\hat{\lambda}_n$ est asymptotiquement négligeable devant la différence $\hat{\lambda}_n - \lambda$.

△

Correction VIII.4. On a $\mathcal{L}(\sum_{i=1}^n X_i) = \Gamma(\lambda, n)$. On en déduit :

$$\mathbb{E}_\lambda[\hat{\lambda}_n] = n \int \frac{\lambda^n}{\Gamma(n)} x^{n-2} \, \mathrm{e}^{-\lambda x} \, \mathbf{1}_{\{x \geq 0\}} \, dx = n\lambda \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{n}{n-1}\lambda.$$

L'estimateur $\dfrac{n-1}{n}\hat{\lambda}_n$ est donc sans biais. Comme la statistique $\bar{X}_n$ est exhaustive et totale (il s'agit de la statistique canonique d'un modèle exponentiel), on en déduit que $\dfrac{n-1}{n}\hat{\lambda}_n$ est un estimateur optimal de λ. La variance de l'estimateur optimal sans biais est :

$$\mathbb{E}_\lambda \left[\left(\frac{n-1}{n}\hat{\lambda}_n - \lambda \right)^2 \right] = \frac{\lambda^2}{n-2} > \frac{1}{nI(\lambda)} \, .$$

L'estimateur $\dfrac{n-1}{n}\hat{\lambda}_n$ n'est donc pas efficace. Enfin on note que $n\left(\frac{n-1}{n}\hat{\lambda}_n - \hat{\lambda}_n\right) = -\hat{\lambda}_n$ converge $\mathbb{P}_\lambda$-p.s., et donc en loi, vers $-\lambda < 0$, alors que $\sqrt{n}\left(\hat{\lambda}_n - \lambda\right)$ converge en loi vers $\mathcal{N}(0, \lambda^2)$. Les normalisations qui assurent une convergence non triviale ne sont pas du même ordre. La différence entre l'estimateur optimal et $\hat{\lambda}_n$ est asymptotiquement négligeable devant la différence $\hat{\lambda}_n - \lambda$. ▲

Exemple VIII.43. Le modèle uniforme $\mathcal{P} = \left\{U_{[0,\theta]}; \theta > 0\right\}$ n'est pas un modèle régulier. Mais on montre que l'estimateur du maximum de vraisemblance converge en $\frac{1}{n}$ (au lieu de $\frac{1}{\sqrt{n}}$ pour les modèles réguliers). On rappelle que l'estimateur du maximum de vraisemblance $\hat{\theta}_n = \max_{1 \leq i \leq n} X_i$ a pour densité $\dfrac{n}{\theta^n} t^{n-1} \mathbf{1}_{[0,\theta]}(t)$. La suite $(n(\theta - \hat{\theta}_n), n \in \mathbb{N}^*)$ converge en loi vers la loi exponentielle de paramètre θ^{-1}. En effet, soit g une fonction mesurable bornée. On a :

$$\mathbb{E}_\theta[g(n(\theta - \hat{\theta}_n))] = \int_0^\theta g(n(\theta - t)) \, \frac{n}{\theta^n} t^{n-1} \, dt.$$

On pose $u = n(\theta - t)$,

$$\mathbb{E}_\theta[g(n(\theta - \hat{\theta}_n))] = \int_{\mathbb{R}^+} g(u) \, \frac{1}{\theta} \left(1 - \frac{u}{n\theta}\right)^{n-1} \mathbf{1}_{\{u \leq n\theta\}} \, du.$$

On suppose que $n \geq 2$. On remarque que la fonction $g(u)f_n(u)$ où $f_n(u) = \left(1 - \frac{u}{n\theta}\right)^{n-1} \mathbf{1}_{\{0 \leq u \leq n\theta\}}$ est bornée en valeur absolue par la fonction intégrable $\|g\|_\infty \, \mathrm{e}^{-u/2\theta} \, \mathbf{1}_{\mathbb{R}^+}(u)$. De plus la suite de fonctions f_n converge simplement vers la fonction $\exp(-u/\theta)\mathbf{1}_{\{u \geq 0\}}$. Par convergence dominée, on obtient que :

$$\mathbb{E}[g(n(\theta - \hat{\theta}_n))] \underset{n\to\infty}{\longrightarrow} \int g(u) \, \frac{1}{\theta} \, \mathrm{e}^{-\frac{u}{\theta}} \, \mathbf{1}_{\{u \geq 0\}} \, du.$$

Cela signifie que $n(\theta - \hat{\theta}_n)$ converge en loi vers la loi exponentielle de paramètre θ^{-1}. La vitesse de convergence est en $\frac{1}{n}$. ◊

Exercice VIII.5.
On conserve les mêmes notations que dans l'exemple VIII.43. On considère l'estimateur sans biais $\hat{\theta}_n^{(1)} = \frac{n+1}{n}\hat{\theta}_n$ et l'estimateur $\hat{\theta}_n^{(2)} = \frac{n+2}{n+1}\hat{\theta}_n$ (préférable aux deux précédents, voir l'exercice VIII.1). Montrer que les suites $(n(\theta - \hat{\theta}_n^{(1)}), n \in \mathbb{N}^*)$ et $(n(\theta - \hat{\theta}_n^{(2)}), n \in \mathbb{N}^*)$ convergent en loi. Identifier les limites. △

VIII.6.4 Comparaison asymptotique

Soit g une fonction définie sur Θ à valeurs réelles et δ_n un estimateur de $g(\theta)$ de carré intégrable convergent et asymptotiquement normal de variance asymptotique $\sigma(\theta)^2$. On suppose de plus que :

$$n\mathbb{E}_\theta\left[(\delta_n - g(\theta))^2\right] \underset{n\to\infty}{\longrightarrow} \sigma(\theta)^2. \tag{VIII.1}$$

Cette égalité est souvent vérifiée, mais elle ne peut pas se déduire directement de la normalité asymptotique. On obtient alors que $\lim_{n\to\infty} nR(\delta_n, \theta) = \sigma(\theta)^2$. La comparaison d'estimateurs à l'aide du risque quadratique pour des estimateurs asymptotiquement normaux se réduit, dans une approche asymptotique, à la comparaison des variances asymptotiques. Cette remarque préliminaire motive la définition suivante.

Définition VIII.44. *Soit deux estimateurs $\delta_n^{(1)}$ et $\delta_n^{(2)}$ de $g(\theta)$ asymptotiquement normaux de variance (ou matrice de covariance) asymptotique $\Sigma_1(\theta)$ et $\Sigma_2(\theta)$. On dit que $\delta_n^{(1)}$ est* ***asymptotiquement préférable*** *à $\delta_n^{(2)}$ si pour tout $\theta \in \Theta$:*

$$\Sigma_1(\theta) \leq \Sigma_2(\theta).$$

Sauf dans des cas triviaux, il n'existe pas d'estimateurs asymptotiquement préférables à tous les autres. En revanche, on peut parfois montrer que les estimateurs du maximum de vraisemblance sont asymptotiquement préférables parmi certaines familles d'estimateurs.

Soit g une fonction réelle définie sur Θ. On déduit du théorème VIII.18 que dans un modèle régulier, si δ_n est un estimateur de $g(\theta)$ sans biais asymptotiquement normal de variance asymptotique $\sigma(\theta)^2$ et si (VIII.1) est vérifiée, alors pour tout $\theta \in \Theta$:

$$\sigma(\theta)^2 \geq \frac{g'(\theta)^2}{I(\theta)}.$$

Cette inégalité est également vraie, sous certaines hypothèses, pour des estimateurs biaisés. Ainsi l'estimateur asymptotiquement normal, dont la variance asymptotique est $g'(\theta)^2/I(\theta)$ est asymptotiquement préférable dans une famille générale d'estimateurs convergents. Plus généralement, on donne la définition suivante.

Définition VIII.45. *Soit g une fonction réelle ou vectorielle définie sur Θ de classe C^1. Un estimateur δ_n de $g(\theta)$ est dit* ***asymptotiquement efficace*** *s'il est asymptotiquement normal de matrice de covariance asymptotique $\Sigma(\theta) = \dfrac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1} \left(\dfrac{\partial g}{\partial \theta}(\theta)\right)^t$.*

Ainsi, un estimateur δ_n de θ asymptotiquement normal de matrice de covariance asymptotique $I(\theta)^{-1}$ est asymptotiquement efficace. Dans le cas réel (θ réel et g réelle), l'estimateur δ_n de $g(\theta)$ est asymptotiquement efficace si :

$$\sqrt{n}(\delta_n - g(\theta)) \xrightarrow[n\to\infty]{\textbf{en loi}} \mathcal{N}\left(0, g'(\theta)^2 I(\theta)^{-1}\right).$$

Exemple. Dans l'exemple VIII.36, l'estimateur $\tilde{\sigma}_n$ est un estimateur sans biais optimal (mais pas efficace) de σ. Il est de plus convergent et asymptotiquement efficace. ◊

Le corollaire suivant est une conséquence directe de la proposition VIII.40.

Corollaire VIII.46. *Soit $\hat{\theta}_n$ un estimateur asymptotiquement efficace de θ. Soit g une fonction réelle ou vectorielle de classe C^1. Alors l'estimateur $g(\hat{\theta}_n)$ est un estimateur asymptotiquement efficace de $g(\theta)$.*

Les estimateurs du maximum de vraisemblance sont dans les cas réguliers des estimateurs asymptotiquement efficaces de θ (cf. théorème VIII.42). La dispersion asymptotique est donc minimale. Cette propriété justifie leur emploi fréquent

dans l'analyse d'échantillons de grande taille. De plus on peut montrer que sous certaines hypothèses, s'il existe un estimateur asymptotiquement efficace de θ, alors le maximum de vraisemblance est un estimateur asymptotiquement efficace.

Même pour les modèles réguliers, il existe des estimateurs asymptotiquement normaux dont la variance asymptotique est inférieure à la matrice de covariance $\Sigma(\theta)$ de la définition VIII.45, avec inégalité stricte pour certains θ. On parle d'estimateur super-efficace. Toutefois leur utilité est limitée car on peut démontrer que l'ensemble des paramètres θ, pour lequel l'inégalité est stricte, est un ensemble très petit.

Exemple. On continue l'exemple VIII.36. L'estimateur $\tilde{\sigma}_n$ de σ dans un modèle gaussien à moyenne connue, est asymptotiquement efficace. On considère un autre estimateur de σ construit à partir de $\frac{1}{n}\sum_{i=1}^n |X_i - \mu_0|$. On a :

$$\mathbb{E}_\sigma\left[\frac{1}{n}\sum_{i=1}^n |X_i-\mu_0|\right] = \sigma\, 2\int_0^\infty x\,\mathrm{e}^{-x^2/2}\,\frac{dx}{\sqrt{2\pi}} = \sigma\sqrt{\frac{2}{\pi}}.$$

L'estimateur $\tilde{\sigma}_n^{(2)} = \sqrt{2\pi}\,\frac{1}{n}\sum_{i=1}^n |X_i-\mu_0|$ de σ est un estimateur sans biais. Par la loi forte des grands nombres il est convergent. Par le théorème central limite, il est asymptotiquement normal. Sa variance est :

$$\mathrm{Var}_\sigma(\hat{\sigma}_n^{(2)}) = \frac{\pi}{2n}\,\mathrm{Var}_\sigma(|X_1|) = \frac{(\pi-2)}{2n}\sigma^2.$$

Cet estimateur n'est pas asymptotiquement efficace (car $I^{-1}(\sigma) = \frac{1}{2}\,\sigma^2 < \frac{(\pi-2)}{2}\sigma^2$). On en déduit que $\hat{\sigma}_n$ lui est asymptotiquement préférable. Il lui est même préférable pour tout $n \geq 2$. ◊

Remarque. Si le modèle n'est pas régulier, l'efficacité asymptotique peut ne pas être pertinente. Voir l'exemple VIII.43. ◊

VIII.7 Résumé

Soit $p(x_1;\theta)$ la densité de X_1 de loi P_θ, et $p_n(x;\theta) = \prod_{k=1}^n p(x_k;\theta)$, où $x = (x_1,\ldots,x_n)$, la densité de l'échantillon $X = (X_1,\ldots,X_n)$ de v.a. indépendantes et de même loi P_θ. Le paramètre $\theta \in \Theta$ est inconnu.

Taille de l'échantillon fixée

– Un estimateur est **préférable** à un autre estimateur s'il diminue la **fonction de risque** pour tout $\theta \in \Theta$.

– Le **biais** d'un estimateur $\hat{g}$ intégrable de $g(\theta)$ est défini par $\mathbb{E}_\theta[\hat{g}] - g(\theta)$. Un estimateur est sans biais si son biais est nul pour tout $\theta \in \Theta$.

– La **vraisemblance** de l'échantillon est la fonction $\theta \mapsto p_n(x;\theta)$; la **log-vraisemblance** $L_n(x;\theta)$ est la fonction de θ définie par $\theta \mapsto \log p_n(x;\theta) = \sum_{k=1}^n \log p(x_k;\theta)$.

– Si g est une bijection et si $\hat{\theta}$ est l'**estimateur du maximum de vraisemblance** de θ, alors $g(\hat{\theta})$ est l'estimateur du maximum de vraisemblance de $g(\theta)$.

– L'estimateur du maximum de vraisemblance est en général biaisé.

– Le **score** de l'échantillon est défini par :

$$V_n = \frac{\partial L_n}{\partial \theta}(X;\theta) = \frac{\partial \log p_n}{\partial \theta}(X;\theta).$$

– S'il est intégrable, alors le score est centré : $\mathbb{E}[V_n] = 0$.

– L'**information de Fisher**, quand elle existe, est définie par :

$$\begin{aligned} I(\theta) = \mathrm{Cov}(V_1, V_1) &= \mathbb{E}\left[\frac{\partial \log p}{\partial \theta}(X_1;\theta)\left(\frac{\partial \log p}{\partial \theta}(X_1;\theta)\right)^t\right] \\ &= -\mathbb{E}\left[\frac{\partial^2 \log p}{\partial \theta^2}(X_1;\theta)\right]. \end{aligned}$$

– L'information de Fisher de l'échantillon de taille n est $I_n(\theta) = nI(\theta)$.

– La fonction de risque quadratique d'un estimateur sans biais de $g(\theta)$, construit à partir d'un échantillon de taille n, est minorée par la **borne FDCR** définie par :

$$\frac{1}{n}\frac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1}\left(\frac{\partial g}{\partial \theta}(\theta)\right)^t.$$

– Si la fonction de risque quadratique d'un estimateur sans biais atteint la borne FDCR, on dit que l'estimateur est **efficace**.

– Soit S une statistique exhaustive, i.e. la loi de l'échantillon sachant S est indépendante du paramètre. Si δ est un estimateur de $g(\theta)$, alors l'estimateur **amélioré de Rao-Blackwell** $\mathbb{E}[\delta|S]$ lui est préférable.

– Un estimateur sans biais de $g(\theta)$ est **optimal** s'il est préférable à tous les estimateurs sans biais.

– Soit S une statistique exhaustive et totale. Si δ est un estimateur sans biais de $g(\theta)$, alors l'estimateur **amélioré de Rao-Blackwell**, $\mathbb{E}[\delta|S]$, est optimal.

Approche asymptotique

Par abus on note δ_n pour la suite d'estimateurs $(\delta_n, n \in \mathbb{N}^*)$ où δ_n est une statistique de l'échantillon de taille n.

- Un estimateur δ_n de $g(\theta)$ est **convergent** si : pour tout $\theta \in \Theta$,

$$\lim_{n\to\infty} \delta_n = g(\theta), \quad \mathbb{P}_\theta\text{-p.s.}$$

- Sous certaines hypothèses, on peut construire une suite convergente d'estimateurs de $g(\theta)$ avec
  - la **méthode de substitution**,
  - la **méthode des moments**,
  - l'**estimateur du maximum de vraisemblance**.
- Une suite d'estimateurs $(\delta_n, n \in \mathbb{N}^*)$ de $g(\theta)$ est **asymptotiquement normale** si : pour tout $\theta \in \Theta$,

$$\sqrt{n}(\delta_n - g(\theta)) \xrightarrow[n\to\infty]{\textbf{en loi}} \mathcal{N}(0, \Sigma(\theta)),$$

  où P_θ est la loi de X_k. La matrice $\Sigma(\theta)$ est la **matrice de covariance asymptotique** de la suite d'estimateurs.
- Sous de bonnes hypothèses, l'estimateur de substitution, l'estimateur des moments et l'estimateur du maximum de vraisemblance sont asymptotiquement normaux.
- Soit $\delta_n^{(1)}$ et $\delta_n^{(2)}$ deux d'estimateurs de $g(\theta)$ asymptotiquement normaux. On dit que $\delta_n^{(1)}$ est **asymptotiquement préférable** à $\delta_n^{(2)}$ si $\Sigma_1(\theta) \leq \Sigma_2(\theta)$ pour tout $\theta \in \Theta$, avec $\Sigma_i(\theta)$ la matrice de covariance asymptotique de $\delta_n^{(i)}$.
- Un estimateur δ_n de $g(\theta)$ est **asymptotiquement efficace** s'il est asymptotiquement normal de matrice de covariance asymptotique :

$$\frac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1} \left(\frac{\partial g}{\partial \theta}(\theta)\right)^t.$$

- Sous de bonnes hypothèses, l'estimateur du maximum de vraisemblance est asymptotiquement efficace.

VIII.8 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice VIII.1 p. 206, *Exercice* VIII.2 p. 215, *Exercice* VIII.3 p. 215, *Exercice* VIII.4 p. 224, *Exercice* VIII.5 p. 226.

Exercice VIII.6.
On considère le modèle d'échantillonnage $(X_1, \dots, X_n)$ de taille n associé à la famille de lois de Poisson $\mathcal{P} = \{\mathcal{P}(\theta), \theta > 0\}$. On cherche à estimer $\mathbb{P}_\theta(X_i = 0)$.

1. Montrer que le modèle est exponentiel. Déterminer la statistique canonique S. Est-elle exhaustive et totale ? Donner sa loi.
2. Calculer $\mathbb{P}_\theta(X_i = 0)$ et montrer que $\mathbf{1}_{\{X_1=0\}}$ en est un estimateur sans biais.
3. Montrer que la loi conditionnelle de X_1 sachant S est une loi binomiale de paramètre $\left(S, \frac{1}{n}\right)$.
4. En déduire que $\delta_S = \left(1 - \frac{1}{n}\right)^S$ est l'estimateur optimal de $\mathbb{P}_\theta(X_i = 0)$. Est-il convergent ?
5. Calculer le score et l'information de Fisher.
6. En déduire la borne FDCR pour l'estimation de $\mathbb{P}_\theta(X_i = 0)$. Est-elle atteinte par δ_S ?
7. L'estimateur δ_S est-il asymptotiquement normal, asymptotiquement efficace ?
8. Calculer l'estimateur du maximum de vraisemblance de $\mathbb{P}_\theta(X_i = 0)$.
9. Vérifier qu'il est biaisé. Est-il préférable à δ_S ?
10. Vérifier qu'il est asymptotiquement efficace.

△

Exercice VIII.7.
On considère la loi de densité $f_\alpha(x) = c_\alpha \frac{1}{x^{\alpha+1}} \mathbf{1}_{\{x \geq 1\}}$ et le modèle d'échantillonnage $(X_1, \dots, X_n)$ de taille n associé.

1. Déterminer les valeurs possibles de α.
2. Calculer c_α.
3. Calculer $\mathbb{E}[X_1]$.
4. En déduire un estimateur convergent de α.
5. À quelle condition sur α, l'estimateur est-il asymptotiquement normal ?
6. L'estimateur est-il asymptotiquement efficace ?

7. Calculer l'estimateur du maximum de vraisemblance de α.
8. Déterminer la loi de $\log(X_1)$.
9. Calculer l'estimateur du maximum de vraisemblance de $1/\alpha$, et donner sa loi.
10. Vérifier directement si cet estimateur est biaisé, efficace et asymptotiquement normal.

△

Exercice VIII.8.
On considère le modèle d'échantillonnage $(X_1, \ldots, X_n)$ de taille n associé à la famille de lois gaussiennes $\mathcal{N}(0, \sigma^2)$, où $\sigma > 0$ est inconnu.

1. Calculer $\mathbb{E}[|X_1|]$.
2. En déduire un estimateur sans biais de σ.
3. Est-il efficace ?
4. Est-il asymptotiquement normal ?
5. Est-il asymptotiquement efficace ?
6. Calculer l'estimateur du maximum de vraisemblance de σ.
7. Cet estimateur est-il biaisé ? En déduire un estimateur non biaisé de σ.
8. Vérifier que ce dernier estimateur est optimal mais pas efficace.
9. Est-il asymptotiquement normal, asymptotiquement efficace ?

△

Exercice VIII.9.
On considère le modèle d'échantillonnage $(X_1, \ldots, X_n)$ de taille n associé à la famille de lois géométriques de paramètre $\theta > 0$.

1. Calculer $\mathbb{E}[X_1]$ et $\mathbb{E}[1/X_1]$.
2. En déduire un estimateur convergent de θ.
3. Calculer l'estimateur du maximum de vraisemblance de θ.
4. Pour un échantillon de taille 1, cet estimateur est-il sans biais ?
5. Vérifier directement si cet estimateur est convergent, asymptotiquement normal et asymptotiquement efficace.

△

Exercice VIII.10.
On considère le modèle d'échantillonnage $(X_1, \ldots, X_n)$ de taille n associé à la famille de lois $\Gamma(\lambda, 2)$, où $\lambda > 0$ est inconnu.

1. Calculer $\mathbb{E}[X_1]$ et $\mathbb{E}[1/X_1]$.

2. En déduire un estimateur convergent et sans biais de λ. Est-il asymptotiquement normal ?
3. Déterminer l'estimateur du maximum de vraisemblance de λ.
4. Pour un échantillon de taille 1, cet estimateur est-il sans biais ?
5. Est-il convergent ?
6. Déterminer la loi de $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ en utilisant les fonctions caractéristiques.
7. En déduire que l'estimateur du maximum de vraisemblance est biaisé.
8. L'estimateur du maximum de vraisemblance est-il asymptotiquement normal ?
9. L'estimateur du maximum de vraisemblance est-il asymptotiquement efficace ?
10. Déduire de l'estimateur du maximum de vraisemblance un estimateur sans biais.
11. Ce dernier estimateur est-il optimal, efficace ?
12. En considérant des estimateurs de la forme $c_n/\bar{X}_n$, trouver un estimateur préférable aux deux précédents estimateurs. Ce dernier estimateur est-il biaisé ?

△

Exercice VIII.11.
On considère le modèle d'échantillonnage $(X_1, \ldots, X_n)$ de taille n associé à la famille de lois gaussiennes $\{\mathcal{N}(\mu, 1), \mu \in \mathbb{R}\}$. On note $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ la moyenne empirique. Soit T un estimateur de carré intégrable de μ.

1. Vérifier que $\bar{X}_n$ est un estimateur efficace de μ.
2. En déduire que si T est un estimateur préférable à $\bar{X}_n$, alors il est biaisé. On note $b(\mu) = \mathbb{E}_\mu[T] - \mu$ son biais.
3. On rappelle que T peut s'écrire $T = h(X_1, \ldots, X_n)$. Montrer que la fonction $\mu \mapsto \mathbb{E}_\mu[T]$ est dérivable de dérivée :

$$-\int_{\mathbb{R}^n} h(x_1, \ldots, x_n) \left[\sum_{i=1}^n (x_i - \mu)\right] \mathrm{e}^{-\sum_{i=1}^n (x_i-\mu)^2/2}\ dx_1 \ldots dx_n.$$

4. En déduire que l'estimateur T est régulier et que son biais est une fonction dérivable de μ.
5. En reprenant la démonstration du théorème qui établit la borne FDCR, vérifier la minoration suivante pour le risque quadratique de T, $R(T, \mu) = \mathbb{E}_\mu[(T-\mu)^2]$:

$$R(T, \mu) \geq b(\mu)^2 + \frac{(1 + b'(\mu))^2}{n}.$$

6. On suppose maintenant que T est préférable à $\bar{X}_n$. En déduire l'inégalité :

$$\frac{1}{n} \geq b(\mu)^2 + \frac{(1+b'(\mu))^2}{n}.$$

7. En déduire que la fonction b est décroissante et bornée. Montrer qu'en fait $b(\mu) = 0$.
8. En déduire qu'il n'existe pas d'estimateur préférable à $\bar{X}_n$.

La conclusion de la dernière question n'est plus vraie si l'on considère les lois gaussiennes sur $\mathbb{R}^3$ de moyenne $\mu \in \mathbb{R}^3$ et matrice de covariance l'identité. On peut en fait montrer dans ce cas que $\bar{X}_n$ est inadmissible. △

IX

Tests d'hypothèses

Un test, décrit au paragraphe IX.1, correspond à une procédure de décision binaire à partir de données observées : en général cela revient à dire, dans un modèle paramétrique, si le paramètre *a priori* inconnu appartient à un sous-ensemble donné, hypothèse H_0 dite nulle, ou à son complémentaire, hypothèse H_1 dite alternative. Le paragraphe IX.2 présente les deux erreurs d'un test : rejeter H_0 à tort (erreur de $1^{\text{ère}}$ espèce) ou rejeter H_1 à tort (erreur de $2^{\text{ème}}$ espèce). En général, dans les cas concrets, les deux erreurs n'ont pas la même importance. On utilise le principe de Neyman, paragraphe IX.3 : on majore ***a priori*** l'erreur de $1^{\text{ère}}$ espèce. On peut alors comparer deux tests en comparant l'erreur de $2^{\text{ème}}$ espèce. En pratique, on choisit les hypothèses H_0 et H_1 en tenant compte du principe de Neyman. La p-valeur introduite au paragraphe IX.5 indique l'erreur de $1^{\text{ère}}$ espèce en fonction des données : c'est la probabilité de rejeter H_0 à tort. On retiendra que si la p-valeur est faible, alors on rejette H_0. La procédure de test est décrite page 260.

Dans le paragraphe IX.4, on établit le test optimal quand les deux hypothèses H_0 et H_1 sont simples, i.e. réduites à des singletons. Le paragraphe IX.6 permet d'étendre, dans le cadre des modèles exponentiels, les tests présentés pour des hypothèses simples à des hypothèses composites. Le paragraphe IX.7 présente les modèles de régression linéaire qui sont très couramment utilisés, ainsi que les tests correspondants d'analyse de la variance. Si l'on dispose d'un grand nombre d'observations, on peut alors suivre l'approche asymptotique développée au paragraphe IX.8, en particulier pour les hypothèses implicites (test de Wald, paragraphe IX.8.2) ou explicites (test de Hausman, paragraphe IX.8.3). Dans le cas de variables aléatoires prenant un nombre fini de valeurs, ce dernier test correspond au test du χ^2 empirique qui est développé au paragraphe IX.9 avec deux applications particulières : le test du χ^2 d'adéquation à une loi, paragraphe IX.9.2, et le test du χ^2 d'indépendance, paragraphe IX.9.3. On donne également, dans le para-

graphe IX.10 d'autres tests asymptotiques importants : le test non-paramétrique de Kolmogorov-Smirnov d'adéquation à une loi, paragraphe IX.10.1, le test de comparaison non-paramétrique entre deux échantillons de Kolmogorov-Smirnov, paragraphe IX.10.2, et le test de comparaison paramétrique entre deux échantillons, paragraphe IX.10.3.

IX.1 Tests

On considère un modèle d'échantillonnage paramétrique $X_1, \ldots, X_n$ de taille n (cf. chapitre VIII.1). Les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et de même loi inconnue P, et elles sont à valeurs dans $\mathcal{X}$ ($\mathbb{R}$ ou $\mathbb{R}^d$). On suppose que la loi P appartient à une famille de probabilités $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$. Soit H_0 et H_1 deux sous-ensembles **disjoints** (non vides) de Θ. Par convention H_0 est appelé **hypothèse nulle** et H_1 **hypothèse alternative**. On désire établir une procédure, appelée **test**, qui permette, à la vue d'une réalisation $x_1, \ldots, x_n$ de l'échantillon, de déterminer si le paramètre inconnu θ appartient à H_0 ou si au contraire il appartient à H_1.

Remarque. Dans le chapitre VII.2, on établit un test qui permet au fabricant de composants d'accepter l'hypothèse nulle $H_0 =]0, \theta_*]$ ou de rejeter H_0 au profit de l'hypothèse alternative $H_1 =]\theta_*, \infty[$. ◊

On dit qu'une hypothèse est **simple** si elle est réduite à un singleton. Sinon on dit qu'elle est **composite**. Afin d'insister sur le paramètre θ utilisé, une hypothèse correspondant à un sous-ensemble A de Θ sera parfois aussi notée $\{\theta \in A\}$.

Exemple IX.1. Un laboratoire pharmaceutique expérimente un nouveau médicament. On note θ_0 la valeur connue du taux de guérison de la maladie avec le médicament de référence et θ la valeur inconnue du taux de guérison quand le nouveau médicament est administré. Le laboratoire espère rejeter l'hypothèse simple nulle $H_0 = \{\theta_0\}$ (ou l'hypothèse composite nulle $H_0 = \{\theta \leq \theta_0\}$) et donc accepter l'hypothèse alternative composite $H_1 = \{\theta > \theta_0\}$. ◊

Définition IX.2. *Un **test pur** est une application φ de $\mathcal{X}^n$ dans $\{0, 1\}$. Pour une réalisation $x = (x_1, \ldots, x_n)$ de l'échantillon de taille n, on accepte H_0 (et donc on rejette H_1) si $\varphi(x) = 0$ et on accepte H_1 (et donc on rejette H_0) si $\varphi(x) = 1$. On dit que $W_n = \{x; \varphi(x) = 1\}$ est la **région critique** (ou zone de rejet) du test pur.*

Si la réalisation observée est dans la région critique, on rejette l'hypothèse nulle H_0.

Définition IX.3. *Un **test aléatoire** est une application $\varphi : \mathcal{X}^n \mapsto [0,1]$, et $\varphi(x)$ est la probabilité de rejeter H_0 lorsque l'on observe la réalisation $x = (x_1, \ldots, x_n)$ pour l'échantillon de taille n.*

Ainsi un test pur est un test aléatoire particulier. En pratique l'usage d'un test qui n'est pas pur semble illusoire : comment justifier une décision sur un coup de dé ? L'intérêt des tests aléatoires réside dans l'analyse complète des tests d'hypothèses simples (cf. le paragraphe IX.4) et dans la description des meilleurs tests (meilleur dans un sens que l'on précisera). Une fois ces tests construits, il est facile de voir que sous certaines conditions naturelles, on obtient des tests purs. L'introduction des tests aléatoires est donc un artifice technique.

IX.2 Erreurs

Il convient de distinguer deux types d'erreurs. L'**erreur de 1ère espèce** consiste à refuser H_0 alors qu'elle est vraie. L'**erreur de 2ème espèce** consiste à accepter H_0 alors qu'elle est fausse.

Dans l'exemple IX.1, l'erreur de première espèce revient à dire que le nouveau médicament est plus efficace que le médicament de référence à tort ; l'erreur de deuxième espèce revient à dire que le nouveau médicament est moins efficace que le médicament de référence à tort.

En général on ne peut diminuer simultanément les deux erreurs. Par convention, **on cherche à minimiser en priorité** l'erreur de 1ère espèce. On retiendra que l'on ne veut pas rejeter H_0 à tort (ou bien que l'on veut accepter H_1 à raison). Ceci entraîne une dissymétrie entre l'hypothèse nulle et l'hypothèse alternative, et permet de justifier du choix de H_0 et H_1. Dans l'exemple IX.1, il est important de ne pas commercialiser le nouveau médicament à tort, autrement dit il est important de commercialiser le nouveau médicament que s'il est meilleur que le médicament de référence. L'hypothèse nulle est donc $H_0 = \{\theta = \theta_0\}$ ou $H_0 = \{\theta \leq \theta_0\}$ et l'hypothèse alternative $H_1 = \{\theta > \theta_0\}$. Dans l'exemple développé au chapitre VII.2, il est important de s'assurer que $\theta > \theta_*$. L'hypothèse nulle est donc $H_0 = \{\theta \in]0, \theta^*]\}$.

Pour φ un test aléatoire, on pose $\mathbb{E}_\theta[\varphi] = \mathbb{E}_\theta[\varphi(X_1, \ldots, X_n)]$.

Définition IX.4. *L'erreur de 1ère espèce d'un test φ est définie par :*

$$\mathbb{E}_\theta[\varphi] \quad \text{où} \quad \theta \in H_0$$

et l'erreur de 2ème espèce est définie par :

$$1 - \mathbb{E}_\theta[\varphi] \quad \textit{où} \quad \theta \in H_1.$$

*La **puissance** d'un test φ est la fonction définie sur H_1 par $\theta \mapsto \rho_\varphi(\theta) = \mathbb{E}_\theta[\varphi]$.*

La quantité $1 - \rho_\varphi(\theta)$ représente l'erreur de 2ème espèce.

Pour un test pur de région critique W_n, on a $\varphi = \mathbf{1}_{W_n}$. Par convention on écrira $\mathbb{P}_\theta(W_n)$ pour $\mathbb{E}_\theta[\varphi] = \mathbb{P}_\theta((X_1, \ldots, X_n) \in W_n)$. Pour $\theta \in H_0$, $\mathbb{P}_\theta(W_n)$ est l'erreur de 1ère espèce, i.e. la probabilité que la réalisation observée soit dans la région critique alors que $\theta \in H_0$. Pour $\theta \in H_1$, $\mathbb{P}_\theta(W_n)$ est la puissance du test. L'erreur de 2ème espèce est $1 - \mathbb{P}_\theta(W_n)$ où $\theta \in H_1$, i.e. la probabilité que la réalisation observée ne soit pas dans la région critique alors que $\theta \in H_1$.

Le niveau d'un test est le supremum des erreurs de 1ère espèce.

Définition IX.5. *Le **niveau** d'un test φ est* $\boxed{\alpha_\varphi = \sup_{\theta \in H_0} \mathbb{E}_\theta[\varphi].}$

Pour un test pur de région critique W_n, son niveau s'écrit aussi $\sup_{\theta \in H_0} \mathbb{P}_\theta(W_n)$.

IX.3 Choix d'un test

L'erreur de 1ère espèce est celle que l'on veut contrôler en priorité. On cherche donc en priorité un test de niveau faible. Le **principe de Neyman** ne retient que les tests φ dont le niveau est inférieur à un **seuil α fixé *a priori***. Les valeurs typiques de α sont 5%, 1%,... Ensuite, parmi les tests de niveau inférieur à α, on cherche à minimiser l'erreur de 2ème espèce. On recherche donc **les tests de niveau inférieur à α et de puissance maximale**.

Définition IX.6. *Un test φ est dit **uniformément plus puissant** (UPP) au **seuil** α s'il est de niveau inférieur à α et si pour tout test φ' de niveau inférieur à α, l'erreur de 2ème espèce de φ' est supérieure à celle de φ : $\mathbb{E}_\theta[\varphi] \geq \mathbb{E}_\theta[\varphi']$ pour tout $\theta \in H_1$.*

On peut dans certains cas trouver des tests UPP, comme par exemple pour le test d'une hypothèse simple contre une hypothèse simple, voir le paragraphe IX.4 ainsi que le paragraphe IX.6.

IX.4 Test d'hypothèses simples

On conserve les notations suivantes (cf. paragraphe VIII.1) : $p(x_1;\theta)$, avec $x_1 \in \mathcal{X}$, représente la densité de la loi de X_1, P_θ, et

$$p_n(x;\theta) = p_n(x_1,\dots,x_n;\theta) = p(x_1;\theta)\cdots p(x_n;\theta),$$

avec $x = (x_1,\dots,x_n) \in \mathcal{X}^n$, représente la densité de l'échantillon $X = (X_1,\dots,X_n)$ où les variables aléatoires X_i sont indépendantes de loi P_θ.

On suppose que $H_0 = \{\theta_0\}$ et $H_1 = \{\theta_1\}$ avec $\mathrm{P}_{\theta_0} \neq \mathrm{P}_{\theta_1}$ (i.e. le modèle $\{\mathrm{P}_\theta, \theta \in \{\theta_0,\theta_1\}\}$ est identifiable).

Définition IX.7. *On dit que le test φ est un* ***test de Neyman*** *s'il existe $\kappa > 0$ et $\gamma \in [0,1]$ tel que pour $x \in \mathcal{X}^n$:*

$$\varphi(x) = \begin{cases} 1 & si \quad p_n(x;\theta_1) > \kappa p_n(x;\theta_0), \\ \gamma & si \quad p_n(x;\theta_1) = \kappa p_n(x;\theta_0), \\ 0 & si \quad p_n(x;\theta_1) < \kappa p_n(x;\theta_0). \end{cases}$$

Le test φ est défini à l'aide du **rapport de vraisemblance** $Z_n(x) = \dfrac{p_n(x;\theta_1)}{p_n(x;\theta_0)}$. On note $Z_n = Z_n(X_1,\dots,X_n)$. Le théorème suivant caractérise les tests UPP pour les tests d'une hypothèse simple contre une hypothèse simple.

Théorème IX.8 (Neyman-Pearson)**.** *On a les propriétés suivantes.*

1. *Pour toute valeur de $\alpha \in]0,1[$, un test de Neyman tel que $\mathbb{E}_{\theta_0}[\varphi] = \alpha$ est un test UPP au seuil α.*
2. *Pour toute valeur $\alpha \in]0,1[$, il existe un test de Neyman de niveau α. Soit κ et γ les paramètres de ce test. Alors κ est le quantile d'ordre $1-\alpha$ de la loi de Z_n sous H_0. Soit F la fonction de répartition de Z_n sous H_0. Si $F(\kappa) = 1-\alpha$ on peut choisir $\gamma = 1$ et alors φ est un test pur de région critique $W_n = \{x, Z_n(x) \geq \kappa\}$. Sinon, on a $\gamma = (F(\kappa) - (1-\alpha))/(F(\kappa) - F(\kappa-))$ et $\gamma \in]0,1]$.*

Démonstration. Soit φ un test de Neyman de niveau α associé au nombre κ. Soit h une fonction définie sur $\mathcal{X}^n$ telle que $0 \leq h \leq 1$. On remarque que :

$$\begin{aligned} \text{si} \quad & p_n(x;\theta_1) - \kappa p_n(x;\theta_0) > 0, \quad \text{alors} \quad \varphi(x) - h(x) = 1 - h(x) \geq 0, \\ \text{et si} \quad & p_n(x;\theta_1) - \kappa p_n(x;\theta_0) < 0, \quad \text{alors} \quad \varphi(x) - h(x) = -h(x) \leq 0. \end{aligned}$$

Donc on a :

$$[\varphi(x) - h(x)]\,[p_n(x;\theta_1) - \kappa p_n(x;\theta_0)] \geq 0. \tag{IX.1}$$

En prenant l'espérance dans (IX.1) avec x remplacé par X, il vient :

$$(\mathbb{E}_{\theta_1}[\varphi] - \mathbb{E}_{\theta_1}[h]) \geq \kappa\,(\mathbb{E}_{\theta_0}[\varphi] - \mathbb{E}_{\theta_0}[h]) = \kappa(\alpha - \mathbb{E}_{\theta_0}[h]). \tag{IX.2}$$

Si h est un test de niveau inférieur à α, on a donc $\alpha - \mathbb{E}_\theta[h] \geq 0$. Cela implique $\mathbb{E}_{\theta_1}[\varphi] - \mathbb{E}_{\theta_1}[h] \geq 0$. Ainsi le test φ est plus puissant que le test h. Ceci démontre la propriété 1.

On montre la propriété 2. Soit F la fonction de répartition du rapport de vraisemblance Z_n sous H_0 : $F(z) = \mathbb{P}_{\theta_0}(Z_n \leq z)$. On rappelle que la fonction F est croissante continue à droite. Elle vérifie $\lim_{z\uparrow 0} F(z) = 0$, $\lim_{z\to\infty} F(z) = 1$ et $F(z) - F(z-) = \mathbb{P}_{\theta_0}(Z = z)$, où $F(z-) = \lim_{y\uparrow z} F(y)$.

Soit $\varphi = \mathbf{1}_{\{Z_n > \kappa\}} + \gamma \mathbf{1}_{\{Z_n = \kappa\}}$ le test de Neyman associé aux paramètres κ et γ définis dans le théorème. On a :

$$F(\kappa-) \leq 1 - \alpha \leq F(\kappa),$$

où la première inégalité découle de la définition III.3 des quantiles et la seconde de la proposition III.7. On en déduit que $\gamma \in [0,1]$. Le niveau du test φ est donné par :

$$\begin{aligned}\mathbb{E}_{\theta_0}[\varphi] &= \mathbb{P}_{\theta_0}\,(Z_n > \kappa) + \gamma \mathbb{P}_{\theta_0}\,(Z_n = \kappa) \\ &= 1 - F(\kappa) + \gamma\big(F(\kappa) - F(\kappa-)\big) \\ &= \alpha.\end{aligned}$$

Le test φ est donc de niveau α. □

Remarque IX.9. En prenant $h = \alpha$ dans (IX.2), on obtient que si φ est un test de Neyman de niveau α alors $\mathbb{E}_{\theta_1}[\varphi] \geq \alpha$. ◊

Remarque IX.10. On peut montrer que s'il existe un test UPP au seuil (ou de niveau) α, alors c'est essentiellement un test de Neyman.

En effet, on considère le test de Neyman φ de niveau α défini en 2 du théorème précédent. Soit h un test UPP au seuil α. Par définition des tests UPP, on a $\mathbb{E}_{\theta_1}[h] \geq \mathbb{E}_{\theta_1}[\varphi]$. L'intégrale en x de la quantité positive (IX.1) est donc nulle. Cela implique que l'inégalité (IX.1) est donc une égalité presque partout. Donc les tests φ et h coïncident presque partout sur $\{x \in \mathcal{X}^n; p_n(x;\theta_1) \neq \kappa p_n(x;\theta_0)\}$. ◊

En conclusion, on retiendra que les seuls tests UPP au seuil α sont essentiellement des tests de Neyman de niveau α. En pratique, si F, la fonction de répartition

du rapport de vraisemblance, Z_n, sous $\mathbb{P}_{\theta_0}$, est continue (par exemple si Z_n est une v.a. continue) on peut construire un test pur UPP au seuil α. Si F n'est pas continue, on peut choisir α judicieusement (de sorte que le quantile d'ordre $1-\alpha$, κ, vérifie $F(\kappa) = 1-\alpha$ et donc $\gamma = 1$) pour obtenir un test pur UPP de seuil α. Dans ces cas, si κ est le quantile d'ordre $1-\alpha$ de la loi du rapport de vraisemblance, la région critique du test pur est :

$$W_n = \{x = (x_1, \ldots, x_n) \in \mathcal{X}^n; p_n(x;\theta_1) \geq \kappa p_n(x;\theta_0)\}.$$

Exemple IX.11. Soit un modèle d'échantillonnage gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2), \mu \in \mathbb{R}\}$ à variance connue. On choisit les hypothèses $H_0 = \{\mu_0\}$ et $H_1 = \{\mu_1\}$, avec $\mu_0 < \mu_1$. Le rapport de vraisemblance est :

$$\begin{aligned} Z_n &= \exp\left(-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (X_i-\mu_1)^2 - \sum_{i=1}^n (X_i-\mu_0)^2\right]\right) \\ &= \exp\left(\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^n X_i\right) \exp -\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2). \end{aligned}$$

La variable aléatoire Z_n est continue sous $\mathbb{P}_{\mu_0}$. La région critique du **test pur** UPP au seuil α est :

$$\begin{aligned} W_n &= \left\{(x_1,\ldots,x_n); \exp\left(-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (x_i-\mu_1)^2 - \sum_{i=1}^n (x_i-\mu_0)^2\right]\right) \geq \kappa\right\} \\ &= \left\{(x_1,\ldots,x_n); \exp\left(\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^n x_i\right) \exp -\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2) \geq \kappa\right\} \\ &= \left\{(x_1,\ldots,x_n); \frac{1}{n}\sum_{i=1}^n x_i \geq c\right\}. \end{aligned}$$

On cherche à déterminer la constante c plutôt que κ. Pour cela, on utilise le fait que le test pur $\mathbf{1}_{W_n}$ est de seuil α :

$$\alpha = \mathbb{P}_{\mu_0}(W_n) = \mathbb{P}_{\mu_0}\left(\bar{X}_n \geq c\right),$$

où $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est la moyenne empirique. Or la loi de $\bar{X}_n$ sous $\mathbb{P}_{\mu_0}$ est la loi gaussienne $\mathcal{N}\left(\mu_0, \frac{\sigma_0^2}{n}\right)$. Donc on a :

$$\alpha = \mathbb{P}\left(\frac{\sigma_0}{\sqrt{n}}Y + \mu_0 \geq c\right) \quad \text{où} \quad \mathcal{L}(Y) = \mathcal{N}(0,1).$$

On en déduit que $c = \mu_0 + \phi_{1-\alpha}\dfrac{\sigma_0}{\sqrt{n}}$ où $\phi_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi gaussienne $\mathcal{N}(0,1)$.

Soit $x = (x_1, \ldots, x_n)$ le résultat de l'observation de l'échantillon. Si $x \in W_n$ c'est-à-dire si $\frac{1}{n}\sum_{i=1}^n x_i \geq c$, alors on rejette l'hypothèse H_0 au profit de son alternative H_1 au niveau α.

On remarquera que la valeur de μ_1 n'intervient pas pour déterminer la région critique. En fait elle intervient dans le calcul de l'erreur de 2$^{\text{ème}}$ espèce. On a :

$$1 - \mathbb{E}_{\mu_1}[\varphi] = \int_{\frac{\sqrt{n}}{\sigma_0}(\mu_1-\mu_0)-\phi_{1-\alpha}}^{+\infty} \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}}.$$

On remarque que l'erreur de 2$^{\text{ème}}$ espèce est une fonction décroissante en n. Intuitivement, plus on a d'observations et moins on a de chance de se tromper. De même, elle est décroissante en μ_1. Plus μ_1 est proche de μ_0, plus l'erreur de 2$^{\text{ème}}$ espèce est grande. ◊

IX.5 Statistique de test et *p*-valeur

On suppose que la région critique d'un test pur construit à partir d'un échantillon de taille n s'écrit $W_n = \{x \in \mathcal{X}^n; \zeta_n(x) \geq c\}$ ou plus généralement $W_n = \{x \in \mathcal{X}^n; g(\zeta_n(x)) \geq c\}$ (en abrégé $W_n = \{\zeta_n \geq c\}$ ou $W_n = \{g(\zeta_n) \geq c\}$) avec c dépendant du seuil et g une fonction mesurable à valeurs réelles (souvent $g = Id$, $g = -Id$ ou $g = |\cdot|$ avec Id la fonction identité). Si de plus, $\zeta_n = \zeta_n(X_1, \ldots, X_n)$ est une statistique, alors on dit que la v.a. ζ_n est une **statistique de test**. En général on choisit les statistiques de test ζ_n de sorte que les régions critiques soient simples, par exemple de forme unilatérale $W_n = \{\zeta_n \geq c_n\}$ ou $W_n = \{\zeta_n \leq c_n\}$ ou de forme bilatérale $W_n = \{a_n \leq \zeta_n \leq b_n\}$.

Définition IX.12. *Soit ζ_n une statistique de test à valeurs dans $\mathbb{R}^d$. Soit g une fonction mesurable définie sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$. On suppose que pour tout $\alpha \in]0,1[$, il existe $c_\alpha \in \mathbb{R}$ tel que la région critique $\{g(\zeta_n) \geq c_\alpha\}$ définit un test pur de niveau α : $\sup_{\theta\in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq c_\alpha) = \alpha$. La p-**valeur** du test est le plus petit niveau pour lequel on rejette H_0 pour les données observées $x^{obs} \in \mathcal{X}^n$:*

$$p\text{-}val = \inf\{\alpha, g(\zeta_n^{obs}) \geq c_\alpha\}, \quad \text{où} \quad \zeta_n^{obs} = \zeta_n(x^{obs}).$$

Le corollaire suivant est immédiat.

Corollaire IX.13. *Sous les hypothèses de la définition IX.12, on a :*

$$p\text{-}val = \sup_{\theta\in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq g(\zeta_n^{obs})), \quad où \quad \zeta_n^{obs} = \zeta_n(x^{obs}). \tag{IX.3}$$

On rejette H_0 au seuil α si et seulement si la p-valeur est inférieure ou égale à α.

Démonstration. On a par définition :

$$p\text{-val} = \inf\{ \sup_{\theta\in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq c_\alpha), g(\zeta_n^{\text{obs}}) \geq c_\alpha\} = \sup_{\theta\in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq g(\zeta_n^{\text{obs}})).$$

Si on rejette H_0 au seuil α alors $g(\zeta_n(x^{\text{obs}})) \geq c_\alpha$ et donc p-val $\leq \alpha$. Si p-val $< \alpha$ alors $g(\zeta_n(x^{\text{obs}})) > c_\alpha$ et donc on rejette H_0 au niveau α. Si p-val $= \alpha$, alors la région critique $\{g(\zeta_n) \geq g(\zeta_n^{\text{obs}})\}$ est de niveau α ; elle contient $g(\zeta_n^{\text{obs}})$. Donc on rejette H_0. □

La p-valeur permet de mieux apprécier le risque pris en rejetant H_0. Intuitivement, la p-valeur est la probabilité que la statistique de test prenne des valeurs "pires" que celles observées. Elle correspond au risque de rejeter H_0 à tort.

Dans l'exemple IX.11, la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est une statistique de test. La p-valeur est donnée par $\mathbb{P}_{\mu_0}(\bar{X}_n \geq \bar{x}_n^{\text{obs}})$, où $\bar{x}_n^{\text{obs}} = \frac{1}{n}\sum_{i=1}^n x_i$ est la moyenne empirique observée. On remarque que la p-valeur est une fonction des observations ; c'est donc la réalisation d'une variable aléatoire. Si on note F la fonction de répartition de $\bar{X}_n$ sous H_0, alors la p-valeur est une réalisation de $1 - F(\bar{X}_n)$. Dans cet exemple, la fonction F est bijective. On déduit donc de la proposition III.22 que $F(\bar{X}_n)$, est de loi uniforme sur $[0, 1]$. Ceci implique que la p-valeur est, sous H_0, la réalisation d'une variable aléatoire uniforme sur $[0, 1]$. On peut vérifier que sous H_1, la p-valeur n'est pas uniforme sur $[0, 1]$ mais a tendance à prendre des valeurs proches de 0. Ce comportement est assez général.

Exemple IX.14. Le résultat de l'exemple IX.11 se généralise aux familles exponentielles définies au paragraphe VIII.5.3 avec paramètre unidimensionnel. On considère le modèle exponentiel de densité :

$$p(x_1, \theta) = h(x_1)C(\theta)\, \mathrm{e}^{Q(\theta)S(x_1)}\,.$$

Le rapport de vraisemblance de l'échantillon de taille n est :

$$Z(x) = \frac{p_n(x;\theta_1)}{p_n(x;\theta_0)} = \frac{C(\theta_1)}{C(\theta_0)}\, \mathrm{e}^{\left(Q(\theta_1)-Q(\theta_0)\right)\sum_{i=1}^n S(x_i)}\,.$$

On suppose que $\boxed{Q(\theta_1) > Q(\theta_0).}$ Un test UPP de niveau α pour accepter ou rejeter l'hypothèse nulle $H_0 = \{\theta_0\}$ contre l'hypothèse alternative $H_1 = \{\theta_1\}$ est le test de Neyman défini par :

$$\varphi(x) = \begin{cases} 1 & \text{si} \quad \zeta_n(x) > c, \\ \gamma & \text{si} \quad \zeta_n(x) = c, \\ 0 & \text{si} \quad \zeta_n(x) < c, \end{cases}$$

où la statistique de test est $\zeta_n(x) = \sum_{i=1}^{n} S(x_i)$. Cette statistique est appelée **statistique canonique** du modèle exponentiel. Les constantes c et γ sont définies par la condition $\mathbb{E}_{\theta_0}[\varphi] = \alpha$. Si la fonction de répartition de ζ_n sous $\mathbb{P}_{\theta_0}$ est continue ou si on choisit α de sorte qu'il existe c tel que $\mathbb{P}_{\theta_0}(\zeta_n \geq c) = \alpha$, alors on peut choisir $\gamma = 1$. Ainsi, si $\gamma = 1$, on rejettera H_0 si l'on observe les données x^{obs} et que $\zeta_n^{\text{obs}} = \zeta_n(x^{\text{obs}}) \geq c$. Enfin si $\zeta_n^{\text{obs}} < c$, on accepte H_0. La p-valeur de ce test est définie par $\mathbb{P}_{\theta_0}(\zeta_n \geq \zeta_n^{\text{obs}})$. ◊

IX.6 Hypothèses composites pour les modèles exponentiels

On suppose que la famille $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ est exponentielle de densité :

$$p(x_1; \theta) = C(\theta) h(x_1)\, \mathrm{e}^{Q(\theta) S(x_1)},$$

où $x_1 \in \mathcal{X}$ et Θ est un sous-ensemble de $\mathbb{R}$, et que la fonction réelle Q est strictement monotone. Quitte à changer S en $-S$, on peut supposer que la fonction Q est **strictement croissante**. On considère les hypothèses **unilatérales** non vides $H_0 = \{\theta \in \Theta; \theta \leq \theta_0\}$ et $H_1 = \{\theta \in \Theta; \theta > \theta_0\}$. On considère la statistique de test canonique $\zeta_n(x) = \sum_{i=1}^{n} S(x_i)$, où $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$.

Proposition IX.15. *Soit $\alpha \in]0,1[$. Il existe un test UPP de niveau α pour tester $H_0 = \{\theta \leq \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$ défini par :*

$$\varphi(x) = \begin{cases} 1 & \textit{si} \quad \zeta_n(x) > c, \\ \gamma & \textit{si} \quad \zeta_n(x) = c, \\ 0 & \textit{si} \quad \zeta_n(x) < c, \end{cases}$$

où c est le quantile d'ordre $1-\alpha$ de ζ_n sous $\mathbb{P}_{\theta_0}$ et γ est déterminé par la condition $\mathbb{E}_{\theta_0}[\varphi] = \alpha$. De plus, la fonction $\mathbb{E}_\theta[\varphi]$ est croissante en θ.

Démonstration. Soit $\theta_1 \geq \theta_0$. On déduit de l'exemple IX.14 que le test φ défini dans la proposition est un test UPP de niveau α pour tester $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta = \theta_1\}$. Soit $\theta_0' < \theta_1'$. De par la définition de c et γ, le test φ est aussi un test UPP pour tester $H_0 = \{\theta = \theta_0'\}$ contre $H_1 = \{\theta = \theta_1'\}$ avec le seuil $\alpha' = \mathbb{E}_{\theta_0'}[\varphi]$. On déduit de la remarque IX.9, que $\mathbb{E}_{\theta_1'}[\varphi] \geq \alpha' = \mathbb{E}_{\theta_0'}[\varphi]$. En particulier, la fonction $\theta \mapsto \mathbb{E}_\theta[\varphi]$ est croissante. Et pour $\theta \leq \theta_0$, on a $\mathbb{E}_\theta[\varphi] \leq \alpha$.

L'ensemble des tests aléatoires :

$$\mathcal{C} = \{\varphi'; \mathbb{E}_\theta[\varphi'] \leq \alpha, \text{ pour tout } \theta \leq \theta_0\}$$

est un sous-ensemble de $\{\varphi'; \mathbb{E}_{\theta_0}[\varphi'] \leq \alpha\}$. Comme φ maximise la puissance, $\mathbb{E}_{\theta_1}[\varphi]$, dans ce dernier ensemble, il maximise aussi la puissance dans $\mathcal{C}$. Ainsi φ est un test UPP pour tester $H_0 = \{\theta \leq \theta_0\}$ contre $H_1 = \{\theta = \theta_1\}$, où $\theta_1 > \theta_0$. Ceci reste vrai pour tout $\theta_1 > \theta_0$ car la définition de φ ne fait intervenir que θ_0. En conclusion, φ est un test UPP pour tester $H_0 = \{\theta \leq \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$. □

L'allure de la courbe $\theta \mapsto \mathbb{E}_\theta[\varphi]$ pour un test UPP est donnée dans la figure IX.1.

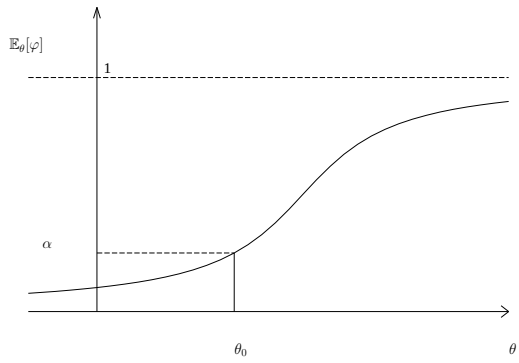


Figure IX.1. Allure typique de la courbe $\theta \mapsto \mathbb{E}_\theta[\varphi]$ pour des tests UPP de seuil α avec $H_0 = \{\theta \leq \theta_0\}$ et $H_1 = \{\theta > \theta_0\}$.

En utilisant la croissance de la fonction $\theta \mapsto \mathbb{E}_\theta[\varphi]$, on déduit le corollaire suivant de la proposition IX.15.

Corollaire IX.16. *On reprend les notations de la proposition IX.15. On suppose que la fonction de répartition F de ζ_n sous $\mathbb{P}_{\theta_0}$ est continue. Alors le test pur de région critique $W_n = \{\zeta_n \geq c\}$ est UPP de niveau α et la p-valeur du test est $\mathbb{P}_{\theta_0}(\zeta_n \geq \zeta_n^{obs}) = 1 - F(\zeta_n^{obs})$, où ζ_n^{obs} est la statistique de test calculée avec les observations. De plus sous $\mathbb{P}_{\theta_0}$ la p-valeur est la réalisation d'une variable aléatoire de loi uniforme sur $[0,1]$.*

Il n'existe pas en général de test UPP pour tester $H_0 = \{\theta = \theta_0\}$ contre l'hypothèse alternative **bilatérale** $H_1 = \{\theta \neq \theta_0\}$. En effet, on considère par exemple le modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$, l'hypothèse nulle $H_0 = \{\mu = \mu_0\}$ et l'hypothèse alternative bilatérale $H_1 = \{\mu \neq \mu_0\}$. On suppose qu'il existe un test UPP φ au seuil α. Alors φ est un test UPP pour tester l'hypothèse nulle $\{\mu = \mu_0\}$ contre l'hypothèse alternative $\{\mu = \mu_1\}$ où $\mu_0 < \mu_1$. D'après la remarque IX.10, il s'agit d'un test de Neyman. Dans ce cas particulier c'est même un test pur de région critique $\{\sum_{i=1}^n X_i \geq c\}$. Mais le test φ est également un test UPP pour l'hypothèse nulle $H_0 = \{\mu = \mu_0\}$ contre l'hypothèse alternative $H_1 = \{\mu = \mu_1\}$ où $\mu_1 < \mu_0$. Il s'agit donc d'un test pur de région critique $\{\sum_{i=1}^n X_i \leq c'\}$. Ceci est absurde. On en déduit donc qu'il n'existe pas de test UPP pour des hypothèses bilatérales.

On choisit de restreindre la classe des tests à la classe des tests sans biais afin de trouver dans cette classe restreinte des tests optimaux.

Définition IX.17. *Un test φ est un test **sans biais** au seuil α si :*

$$\boxed{\sup_{\theta \in H_0} \mathbb{E}_\theta[\varphi] \leq \alpha} \quad \textit{et} \quad \boxed{\inf_{\theta \in H_1} \mathbb{E}_\theta[\varphi] \geq \alpha.}$$

*Un test φ est **uniformément plus puissant sans biais** (UPPS) au seuil α si pour tout test φ' sans biais au seuil α, on a $\mathbb{E}_\theta[\varphi] \geq \mathbb{E}_\theta[\varphi']$ pour tout $\theta \in H_1$.*

On considère toujours les familles exponentielles où la fonction Q est **strictement croissante**. On suppose que $[\theta_1, \theta_2] \subset \Theta$, où Θ est un ouvert de $\mathbb{R}$. On admet le théorème suivant.

Théorème IX.18. *Il existe un test UPPS au seuil α pour tester $H_0 = [\theta_1, \theta_2]$ contre $H_1 = \Theta \backslash [\theta_1, \theta_2]$. Le test est défini par :*

$$\varphi(x) = \begin{cases} 1 & si \quad \zeta_n(x) \notin [c_1, c_2], \\ \gamma_i & si \quad \zeta_n(x) = c_i, \quad où \quad i = 1, 2, \\ 0 & si \quad c_1 < \zeta_n(x) < c_2. \end{cases}$$

Les constantes c_i et γ_i sont déterminées par les conditions :

$$\mathbb{E}_{\theta_1}[\varphi] = \mathbb{E}_{\theta_2}[\varphi] = \alpha.$$

De plus, le test minimise $\mathbb{E}_\theta[\varphi]$ pour $\theta \in [\theta_1, \theta_2]$ et maximise la puissance $\mathbb{E}_\theta[\varphi]$ pour $\theta \in \Theta \backslash [\theta_1, \theta_2]$.

L'allure de la courbe $\theta \mapsto \mathbb{E}_\theta[\varphi]$ pour un test UPPS est donnée dans la figure IX.2.

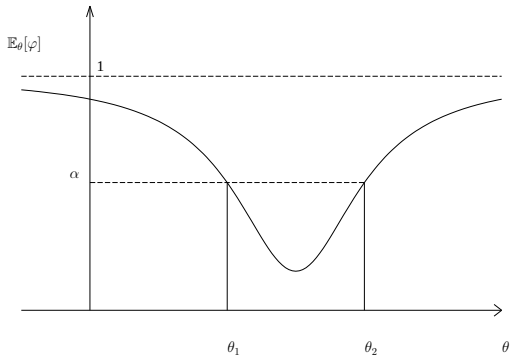


Figure IX.2. Allure typique de la courbe $\theta \mapsto \mathbb{E}_\theta[\varphi]$ pour des tests UPPS de seuil α avec $H_0 = \{\theta \in [\theta_1, \theta_2]\}$ et $H_1 = \{\theta \notin [\theta_1, \theta_2]\}$ et $\theta_1 < \theta_2$.

Remarque. Si la fonction de répartition de la loi de ζ_n est continue sous $\mathbb{P}_{\theta_i}$, $i = 1, 2$, alors on peut choisir $\gamma_i = 1$. On obtient donc pour le test UPPS un test pur de région critique $W_n = \{x; \zeta_n(x) \notin]c_1, c_2[\}$. ◊

Exemple IX.19. On considère le modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$, l'hypothèse nulle $H_0 = \{\mu_0\}$ et l'hypothèse alternative $H_1 = \mathbb{R}\backslash\{\mu_0\}$. La statistique de test est $\zeta_n(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$. Le test UPPS de seuil α est un test pur de région critique $W_n = \{x; \zeta_n(x) \notin]c_1, c_2[\}$. Dans le cas où, avec les notations du théorème, $\theta_1 = \theta_2 = \mu_0$, la seule équation $\mathbb{P}_{\mu_0}(W_n) = \alpha$ ne permet pas de déterminer c_1 et c_2. La puissance du test est :

$$\rho(\mu) = \mathbb{P}_\mu(\zeta_n \notin]c_1, c_2[) = 1 - \int_{\sqrt{n}(c_1-\mu)/\sigma_0}^{\sqrt{n}(c_2-\mu)/\sigma_0} \mathrm{e}^{-y^2/2} \ \frac{dy}{\sqrt{2\pi}}.$$

Elle est minimale par définition pour $\mu = \mu_0$ et de plus $\rho(\mu_0) = \alpha$. Les constantes c_1 et c_2 sont donc déterminées par $\rho'(\mu_0) = 0$ et $\rho(\mu_0) = \alpha$. La condition $\rho'(\mu_0) = 0$ implique $-(c_1 - \mu_0) = c_2 - \mu_0$. Comme $\rho(\mu_0) = \alpha$, on en déduit que $c_1 = \mu_0 - \phi_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$ et $c_2 = \mu_0 + \phi_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$, où $\phi_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi gaussienne $\mathcal{N}(0,1)$. Ainsi le test UPPS de seuil 5% ($\phi_{1-\alpha/2} \simeq 1,96$) est le test pur de région critique :

$$W_n = \left\{ x; \zeta_n(x) \notin \left]\mu_0 \pm \frac{1,96\ \sigma_0}{\sqrt{n}}\right[\right\}.$$

On peut alors calculer la p-valeur associée à ce test pur et aux observations x^{obs} : $p\text{-val} = \mathbb{P}_{\mu_0}(\zeta_n \notin]-\zeta_n^{\text{obs}}, \zeta_n^{\text{obs}}[) = \mathbb{P}_{\mu_0}(|\zeta_n| \geq |\zeta_n^{\text{obs}}|)$, où $\zeta_n^{\text{obs}} = \zeta_n(x^{\text{obs}})$. ◊

IX.7 Régression linéaire

Les modèles de régression linéaire permettent de rendre compte d'un phénomène aléatoire comme combinaison linéaire ou affine de variables explicatives appelées aussi régresseurs. Par exemple, on peut modéliser la taille d'une personne par une combinaison linéaire de la taille de ses deux parents. Les modèles de régression linéaire sont très couramment utilisés. Dans le paragraphe IX.7.1 on détaille l'estimation des coefficients de la régression linéaire et dans le paragraphe IX.7.2 on propose un test pour la nullité de certains coefficients de la régression linéaire.

IX.7.1 Modèle et estimation

On note X_i le phénomène aléatoire mesuré sur l'expérience (ou l'individu) $i \in \{1, \ldots, n\}$, $R_i^1, \ldots, R_i^p$ les valeurs des p régresseurs déterministes pour l'expérience i. On suppose que le phénomène aléatoire s'exprime comme la somme d'une combinaison linéaire ou affine des régresseurs et d'un bruit, qui modélise l'aléa ou les effets de variables explicatives non modélisées. Les bruits sont supposés indépendants gaussiens centrés et de même variance. Plus précisément le modèle s'écrit :

$$X_i = \beta_0 + \sum_{k=1}^{p} \beta_k R_i^k + \varepsilon_i, \quad i \in \{1, \ldots, n\}, \tag{IX.4}$$

où $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^t$ est un vecteur gaussien $\mathcal{N}(0, \sigma^2 I_n)$, avec I_n la matrice identité de taille $n \times n$. Le paramètre du modèle est $\theta = (\beta, \sigma^2)$ où $\beta = (\beta_0, \ldots, \beta_p)^t \in \mathbb{R}^{p+1}$ et $\sigma \in]0, \infty[$. Le modèle (IX.4) peut se réécrire sous forme matricielle :

$$X = M\beta + \varepsilon,$$

où M est la matrice de taille $n \times (p+1)$:

$$M = \begin{pmatrix} 1 & R_1^1 & \dots & R_1^p \\ \vdots & \vdots & & \vdots \\ 1 & R_n^1 & \dots & R_n^p \end{pmatrix} = (\mathbf{1}_n, R^1, \dots, R^p)$$

et $\mathbf{1}_n \in \mathbb{R}^n$ est le vecteur colonne dont tous les coefficients sont égaux à 1. On rappelle que $|\cdot|$ désigne la norme euclidienne sur $\mathbb{R}^n$. Pour un vecteur x et un sous-espace vectoriel E, on note x_E la projection orthogonale de x sur E. La proposition suivante donne la loi de l'estimateur du maximum de vraisemblance de θ. Soit $E = \{Mu, u \in \mathbb{R}^{p+1}\}$ le sous-espace vectoriel de $\mathbb{R}^n$ de dimension $p+1$ engendré par $\mathbf{1}_n$ et les p régresseurs.

Proposition IX.20. *On suppose que la matrice M de taille $n \times (p+1)$ est de rang $p+1$ (i.e. aucun régresseur n'est une combinaison linéaire de $\mathbf{1}_n$ et des autres).*

- *L'estimateur du maximum de vraisemblance de β est :*

$$\hat{\beta} = (M^t M)^{-1} M^t X.$$

  Il est sans biais et de loi gaussienne $\mathcal{N}(\beta, \sigma^2 (M^t M)^{-1})$. La projection orthogonale de X sur E est $X_E = M\hat{\beta}$.
- *L'estimateur de σ^2 donné par :*

$$\hat{\sigma}^2 = \frac{|X - M\hat{\beta}|^2}{n - (p+1)}$$

  est sans biais et égal à l'estimateur du maximum de vraisemblance de σ^2 à une constante multiplicative près. De plus $(n - (p+1))\hat{\sigma}^2/\sigma^2$ suit la loi $\chi^2(n - (p+1))$.
- *Les estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.*

L'estimateur $\hat{\sigma}^2$ de σ^2 a été historiquement préféré à l'estimateur du maximum de vraisemblance car il est sans biais. Au sens de la norme euclidienne, $M\hat{\beta}$ est la meilleure approximation de X à l'aide de $\mathbf{1}_n$ et des régresseurs. On parle aussi d'estimateur des moindres carrés.

On appelle vecteur des résidus le vecteur $\hat{\varepsilon} = X - M\hat{\beta}$. La somme des carrés des résidus, $|\hat{\varepsilon}|^2 = |X - M\hat{\beta}|^2$ correspond à l'erreur qui n'est pas expliquée par le modèle.

Démonstration. Par hypothèse $X - M\beta$ est de loi $\mathcal{N}(0, \sigma^2 I_n)$. La vraisemblance du modèle est donc :

$$p_n(x;\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x_i-\beta_0-\sum_{k=1}^{p}\beta_k R_i^k)^2/(2\sigma^2)},$$

où $x = (x_1, \ldots, x_n)$. On calcule l'estimateur du maximum de vraisemblance de θ. La log-vraisemblance est :

$$\begin{aligned} L_n(X;\theta) &= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \beta_0 - \sum_{k=1}^{p}\beta_k R_i^k)^2 \\ &= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\left|X - M\beta\right|^2. \end{aligned}$$

On a $|X - M\beta|^2 = (X - M\beta)^t(X - M\beta) = X^tX - 2\beta^tM^tX + \beta^t(M^tM)\beta$. La matrice M^tM est symétrique définie positive (car M est de rang $p+1$) et donc inversible. La fonction $\beta \mapsto |X - M\beta|^2$ est donc strictement convexe. Son gradient est donné par $-2M^tX + 2M^tM\beta$, qui s'annule uniquement en $\hat{\beta} = (M^tM)^{-1}M^tX$. On a :

$$\frac{\partial L_n(\hat{\beta},\sigma)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{|X - M\hat{\beta}|^2}{\sigma^3}.$$

Cette dérivée s'annule en $\tilde{\sigma}^2 = |X - M\hat{\beta}|^2/n$. Elle est positive à gauche de $\tilde{\sigma}$ et négative à droite. On en déduit que la log-vraisemblance est maximale en $(\hat{\beta}, \tilde{\sigma})$. L'estimateur du maximum de vraisemblance de (β, σ^2) est donc $(\hat{\beta}, \tilde{\sigma}^2)$.

On a $\hat{\beta} = \beta + (M^tM)^{-1}M^t\varepsilon$. On en déduit que $\hat{\beta}$ suit la loi gaussienne centrée de matrice de covariance :

$$(M^tM)^{-1}M^t((M^tM)^{-1}M^t)^t = (M^tM)^{-1}.$$

Comme $|X - M\hat{\beta}|^2 = \inf_{u\in\mathbb{R}^{p+1}} |X - Mu|^2$, on en déduit que la projection de X sur E est $X_E = M\hat{\beta}$. Comme $M\beta \in E$, la projection orthogonale de ε sur E est $\varepsilon_E = X_E - M\beta$. En particulier $X - X_E = \varepsilon - \varepsilon_E$ est la projection de ε sur l'orthogonal de E. On déduit du théorème de Cochran VI.11, que ε_E, et donc X_E, est indépendant de $\varepsilon - \varepsilon_E = X - M\hat{\beta}$. Le vecteur $\hat{\beta} = (M^tM)^{-1}M^tX_E$ est donc indépendant de $\tilde{\sigma}^2 = |X - M\hat{\beta}|^2/n$. De plus la loi de $|X - M\hat{\beta}|^2/\sigma^2 = |\varepsilon - \varepsilon_E|^2/\sigma^2$ est la loi du χ^2 à $n - (p+1)$ degrés de liberté. □

Exercice IX.1.
On conserve les notations de ce paragraphe.

1. Déterminer la loi de $\hat{\beta}_k - \beta_k$ pour $k \in \{0, \ldots, p\}$.
2. En déduire que si $t_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi de Student de paramètre $n - (p+1)$, alors :

$$\left[\hat{\beta}_k \pm \sqrt{\left((M^tM)^{-1}\right)_{k+1,k+1}}\,\hat{\sigma}t_{1-\alpha/2}\right]$$

est un intervalle de confiance de niveau exact $1-\alpha$ pour β_k.

△

IX.7.2 Test d'utilité des régresseurs

On conserve les notations du paragraphe précédent. Le modèle linéaire permet de tester facilement l'utilité de $p-q$ régresseurs ($0 \leq q < p$). Quitte à renuméroter les régresseurs, on supposera que l'on teste l'utilité des $p-q$ derniers régresseurs. Cela correspond à l'hypothèse nulle $H_0 = \{\beta_j = 0, \quad j \in \{q+1, \ldots, p\}\}$ et à l'hypothèse alternative $H_1 = \{\text{il existe } j \in \{q+1, \ldots, p\}, \text{ tel que } \beta_j \neq 0\}$. Sous H_0, le modèle s'écrit $X = M_0\beta^0 + \varepsilon$ où la matrice $M_0 = (\mathbf{1}_n, R^1, \ldots, R^q)$ est de taille $n \times (q+1)$ et $\beta^0 = (\beta_0^0, \ldots, \beta_q^0)^t$.

On note $H = \{M_0 v, v \in \mathbb{R}^{q+1}\}$ l'espace vectoriel engendré par les régresseurs supposés utiles $\mathbf{1}_n, R^1, \ldots, R^q$.

Proposition IX.21. *On suppose que la matrice M de taille $n \times (p+1)$ est de rang $p+1$. Sous H_0, l'estimateur du maximum de vraisemblance de β^0 est $\hat{\beta}^0 = (M_0^tM_0)^{-1}M_0^tX$. La projection orthogonale de X sur H est $X_H = M_0\hat{\beta}^0$.*

Sous H_0, la variable

$$\zeta_n = \frac{|X_E - X_H|^2/(p-q)}{|X - X_E|^2/(n-(p+1))}$$

suit la loi de Fisher-Snedecor de paramètre $(p-q, n-(p+1))$.

Démonstration. La première partie découle directement de la proposition IX.20.

On note E_1 le sous-espace vectoriel orthogonal de E dans $\mathbb{R}^n$, et E_2 le sous-espace vectoriel orthogonal de H dans E. On a ainsi $E = E_2 \oplus H$ et $E_1 \oplus E_2 \oplus H = \mathbb{R}^n$. Comme $M_0\beta^0$ appartient à H, il vient $\varepsilon_{E_1} = X - X_E$ et $\varepsilon_{E_2} = X_E - X_H$. D'après le théorème de Cochran VI.11, les vecteurs aléatoires ε_{E_1} et ε_{E_2} sont indépendants et $|\varepsilon_{E_1}|^2/\sigma^2$ (resp. $|\varepsilon_{E_2}|^2/\sigma^2$) suit la loi du χ^2 de paramètre $n-(p+1)$ (resp. $p-q$). On en déduit que $\varepsilon_{E_2}/\varepsilon_{E_1}$ suit la loi de Fisher-Snedecor de paramètre $(p-q, n-(p+1))$. □

Le comportement sous H_1 est plus délicat à décrire. On reprend les notations de la démonstration ci-dessus. On a $X_E - X_H = M\beta - (M\beta)_H + \varepsilon_{E_2}$ où $(M\beta)_H$ est la projection de $M\beta$ sur H. Le terme $M\beta - (M\beta)_H$ est non nul sous H_1, et donc $|X_E - X_H|$ a tendance à prendre des valeurs plus élevées que les valeurs usuelles d'un χ^2 à $p-q$ degré de liberté. On remarque que $X_E - X_H$ et $X - X_E$ sont toujours indépendants et que la loi de $X - X_E$ reste inchangée. Donc sous H_1 la variable ζ_n a tendance à prendre des valeurs plus élevées. Ceci conduit à choisir la région critique associée à la statistique de test ζ_n de la forme $W_n = \{\zeta_n \geq z\}$. La p-valeur du test est p-val $= \mathbb{P}(F \geq \zeta_n^{\text{obs}})$, où F suit la loi de Fisher-Snedecor de paramètre $(p-q, n-(p+1))$ et ζ_n^{obs} est la statistique ζ_n évaluée sur les observations. En particulier, sous H_0, la p-valeur est la réalisation d'une variable aléatoire de loi uniforme sur $[0,1]$. Sous H_1, la p-valeur est plutôt faible car ζ_n^{obs} a tendance à prendre des valeurs plus élevées. On rejette donc H_0 lorsque la p-valeur est faible.

Les résultats du test de l'utilité des régresseurs sont habituellement résumés dans la **table d'analyse de la variance** IX.1 avec les acronymes anglo-saxons suivants : SS pour *Sum of Squares*, DF pour *Degrees of Freedom*, MSM pour *Mean Squares of the Model* (associé au choix du sous-modèle) et MSE pour *Mean Squares of the Error* (associé aux erreurs dues au modèle) .

SS	DF	MS	Fisher	p-valeur
$\lvert X_E - X_H\rvert^2$	$p-q$	$MSM = \dfrac{\lvert X_E - X_H\rvert^2}{p-q}$	$\zeta_n^{\text{obs}} = \dfrac{MSM}{MSE}$	$\mathbb{P}(F \geq \zeta_n^{\text{obs}})$
$\lvert X - X_E\rvert^2$	$n-(p+1)$	$MSE = \dfrac{\lvert X - X_E\rvert^2}{n-(p+1)}$		
$\lvert X - X_H\rvert^2$	$n-(q+1)$			

Table IX.1. Table d'analyse de la variance.

Exemple IX.22. **La régression linéaire simple**. On dispose d'un seul régresseur ($p=1$), noté R. Le modèle s'écrit :

$$X_i = \beta_0 + \beta_1 R_i + \varepsilon_i, \quad i \in \{1, \dots, n\}.$$

La loi de X_i est $\mathcal{N}(\beta_0 + \beta_1 R_i, \sigma^2)$. On a $M = (\mathbf{1}_n, R)$ et

$$(M^t M)^{-1} = \frac{1}{n\sum_{i=1}^n (R_i - \bar{R})^2} \begin{pmatrix} \sum_{i=1}^n R_i^2 & -\sum_{i=1}^n R_i \\ -\sum_{i=1}^n R_i & n \end{pmatrix},$$

où $\bar{R} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n R_i$. Le calcul de $\hat{\beta}$ donne :

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(R, X)}{\mathrm{Var}(R)} \quad \text{et} \quad \hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{R},$$

avec $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$ et les notations conventionnelles pour la variance et la covariance empirique : $\mathrm{Var}(R) = \frac{1}{n}\sum_{i=1}^{n}(R_i - \bar{R})^2$ et

$$\mathrm{Cov}(R, X) = \frac{1}{n}\sum_{i=1}^{n}(R_i - \bar{R})(X_i - \bar{X}) = \frac{1}{n}\sum_{i=1}^{n} R_i X_i - \bar{R}\bar{X}.$$

Ces estimateurs coïncident avec les estimateurs des moindres carrés de la droite de régression de X sur R, c'est-à-dire les coefficients b_0 et b_1 de la droite d'équation $X = b_0 + b_1 R$ qui minimisent $\sum_{i=1}^{n}(X_i - b_0 - b_1 R_i)^2$. On obtient que :

$$\hat{\beta}_0 \text{ est de loi } \mathcal{N}\left(\beta_0, \frac{\sigma^2 \sum_{i=1}^{n} R_i^2}{n^2 \,\mathrm{Var}(R)}\right) \quad \text{et} \quad \hat{\beta}_1 \text{ est de loi } \mathcal{N}\left(\beta_1, \frac{\sigma^2}{n\,\mathrm{Var}(R)}\right).$$

On a $X_E = \hat{\beta}_0 \mathbf{1}_n + \hat{\beta}_1 R$, et $SSE = |X - X_E|^2 = \sum_{i=1}^{n}(X_i - \hat{\beta}_0 - \hat{\beta}_1 R_i)^2$. La variable SSE mesure l'écart entre les valeurs observées et les valeurs ajustées par le modèle. Elle suit la loi $\sigma^2\chi^2(n-2)$. L'estimateur sans biais de la variance σ^2 est $|X - X_E|^2/(n-2)$; il est indépendant de $(\hat{\beta}_0, \hat{\beta}_1)$. Cette dernière propriété permet de construire des intervalles de confiance pour β_0 et β_1 de niveau exact à l'aide des lois de Student (voir l'exercice IX.1).

Pour tester l'utilité du régresseur R, on considère les hypothèses $H_0 = \{\beta_1 = 0\}$ et $H_1 = \{\beta_1 \neq 0\}$. On a $X_H = \bar{X}\mathbf{1}_n$ et $SSM = |X_E - X_H|^2 = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 R_i - \bar{X})^2$. La statistique de test est :

$$\zeta_n = \frac{|X_E - X_H|^2}{|X - X_E|^2/(n-2)}.$$

Elle suit sous H_0 la loi de Fisher-Snedecor de paramètre $(1, n-2)$.

On rejette donc H_0 au niveau α si la statistique observée, ζ_n^{obs} est supérieure au quantile d'ordre $1-\alpha$ de la loi de Fisher-Snedecor de paramètre $(1, n-2)$. La p-valeur du test est $p\text{-val} = \mathbb{P}(F \geq \zeta_n^{\mathrm{obs}})$, où F suit la loi de Fisher-Snedecor de paramètre $(1, n-2)$. ◊

Exemple IX.23. Le tableau IX.2 donne le nombre de jours de pluie et la hauteur de pluie en mm, observés pendant toute l'année à Paris de 1956 à 1995. Une représentation sur un graphique des données avec en abscisse le nombre de jours de pluie et en ordonnée la hauteur de pluie, voir le graphique gauche de la figure IX.3, permet de constater que l'ensemble des points forme un nuage allongé et que la quantité de pluie augmente lorsque le nombre de jours de pluie augmente. On désire savoir si l'on peut expliquer la hauteur de pluie par une transformation affine du nombre de jours de pluie. Il s'agit donc du modèle décrit dans l'exemple IX.22 avec X_i la hauteur de pluie de l'année $1955 + i$ et R_i le nombre de jours de pluie de cette même année.

On obtient les résultats suivants :

Années	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Jours	154	161	193	131	198	152	159	159	146	196
Hauteur	545	536	783	453	739	541	528	559	521	880
Années	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
Jours	192	161	176	173	199	141	170	156	198	164
Hauteur	834	592	634	618	631	508	740	576	668	658
Années	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Jours	135	179	171	172	170	197	173	177	177	163
Hauteur	417	717	743	729	690	746	700	623	745	501
Années	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jours	176	180	167	140	149	140	154	155	192	162
Hauteur	611	707	734	573	501	472	645	663	699	670

Table IX.2. Jour et quantité de pluie (en mm) par années à Paris.

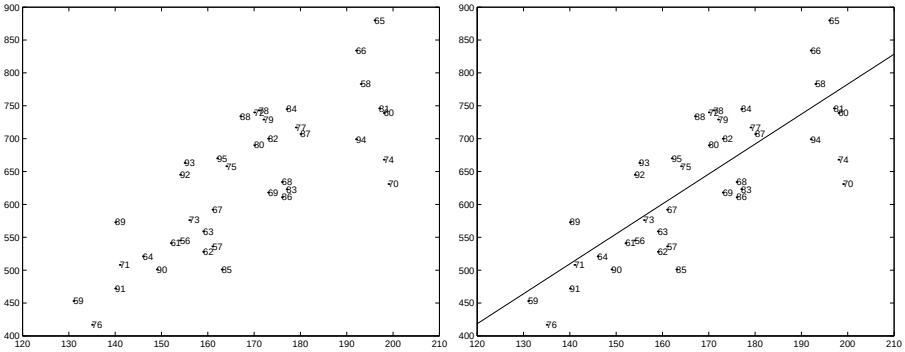


Figure IX.3. Représentation des données (à gauche) et droite de régression sur le nuage de points (à droite).

- Les estimations des paramètres valent : $\hat{\beta}_0 = -128.07$ et $\hat{\beta}_1 = 4.55$. Sur le graphique droit de la figure IX.3 on a représenté la droite de régression.
- Les intervalles de confiance à 95% (donnés dans l'exercice IX.1) sont : $I(\beta_0) = [-322; 66]$ et $I(\beta_1) = [3.40; 5.70]$.
- On obtient la table d'analyse de la variance IX.3 pour le test de $H_0 = \{\beta_1 = 0\}$ contre $H_1 = \{\beta_1 \neq 0\}$. On rejette donc clairement H_0.

SS	DF	MS	Fisher	p-valeur
284135	1	284135	64.5	10^{-9}
167331	38	4403		
451467	39			

Table IX.3. Table d'analyse de la variance.

$\Diamond$

IX.8 Tests asymptotiques

Si l'on dispose d'un grand nombre d'observations, on peut alors utiliser une approche asymptotique. Celle-ci est présentée dans le paragraphe IX.8.1. On donne page 260 le cadre formel de présentation d'un test. On présente deux cas particuliers : le test de Wald pour les hypothèses implicites, paragraphe IX.8.2, et le test de Hausman pour les hypothèses explicites, paragraphe IX.8.3.

IX.8.1 Définitions et exemples

On étudie dans ce paragraphe les comportements asymptotiques de suites de tests.

On considère dorénavant une suite de tests purs $(\mathbf{1}_{W_n}, n \in \mathbb{N}^*)$, où W_n est une région critique correspondant à n observations. Par abus, on confondra test pur, région critique et suite de tests purs.

Définition IX.24. *Un test pur asymptotique $(W_n, n \in \mathbb{N}^*)$ est de **niveau asymptotique** α si :*

$$\boxed{\sup_{\theta \in H_0} \lim_{n\to\infty} \mathbb{P}_\theta(W_n) = \alpha.}$$

*Un test pur asymptotique $(W_n, n \in \mathbb{N}^*)$ est de niveau asymptotique **uniforme** α si :*

$$\boxed{\lim_{n\to\infty} \sup_{\theta \in H_0} \mathbb{P}_\theta(W_n) = \alpha.}$$

*Un test est **convergent** si pour tout $\theta \in H_1$, on a $\lim_{n\to\infty} \mathbb{P}_\theta(W_n) = 1$.*

La convergence du test assure qu'asymptotiquement l'erreur de 2ème espèce est nulle (attention ceci n'est pas uniforme en $\theta \in H_1$). Il est en général difficile de construire des tests purs asymptotiques de niveau asymptotique uniforme fixé.

Dans l'approche asymptotique, on cherche à construire :
- une suite de statistiques de test $(\zeta_n, n \in \mathbb{N}^*)$,
- une fonction mesurable g à valeurs réelles et une suite décroissante $(c_\alpha, \alpha \in]0,1[)$ de réels,

telles que la région critique $\{g(\zeta_n) \geq c_\alpha\}$ est soit de niveau asymptotique uniforme α :

$$\lim_{n\to\infty} \sup_{\theta \in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq c_\alpha) = \alpha, \tag{IX.5}$$

soit de niveau asymptotique non-uniforme α :

$$\sup_{\theta\in H_0} \lim_{n\to\infty} \mathbb{P}_\theta(g(\zeta_n) \geq c_\alpha) = \alpha. \tag{IX.6}$$

Soit x^{obs} une observation de l'échantillon de taille n et $\zeta_n^{\text{obs}} = \zeta_n(x^{\text{obs}})$ la statistique de test évaluée en les observations. En s'inspirant de la définition IX.12 et du corollaire IX.13, on définit la p-valeur asymptotique uniforme si (IX.5) est vérifiée par :

$$p\text{-val} = \lim_{n\to+\infty} \sup_{\theta\in H_0} \mathbb{P}_\theta(g(\zeta_n) \geq g(\zeta_n^{\text{obs}})), \tag{IX.7}$$

et la p-valeur asymptotique non-uniforme si (IX.6) est vérifiée par :

$$p\text{-val} = \sup_{\theta\in H_0} \lim_{n\to+\infty} \mathbb{P}_\theta(g(\zeta_n) \geq g(\zeta_n^{\text{obs}})). \tag{IX.8}$$

Souvent on ne dispose que de la p-valeur non-uniforme. On retiendra que l'on rejette H_0 quand la p-valeur est faible.

On traite en détail dans l'exemple suivant le modèle de Bernoulli avec des hypothèses unilatérales.

Exemple IX.25. On reprend l'exemple IX.1. Un laboratoire pharmaceutique expérimente un nouveau médicament sur des patients. On modélise l'effet du nouveau médicament sur le patient n par une variable aléatoire X_n qui vaut 1 si, utilisant le nouveau médicament, le malade guérit et 0 sinon. On considère un modèle de Bernoulli : $(X_n, n \in \mathbb{N}^*)$ est une suite de variables aléatoires indépendantes de même loi appartenant à $\mathcal{P} = \{\mathrm{P}_\theta, \theta \in]0,1[\}$ où P_θ désigne la loi de Bernoulli de paramètre θ.

Le laboratoire pharmaceutique commercialise le nouveau médicament si son taux de guérison est meilleur que le taux de guérison du médicament de référence, c'est-à-dire si le paramètre θ inconnu de la loi de Bernoulli de X_n est supérieur à θ_0 le taux de guérison connu du médicament de référence. Il dispose de n observations (guérison/non guérison) sur des patients. Il ne peut se permettre de mettre sur le marché un nouveau médicament qui soit moins efficace que le médicament de référence. Il ne peut donc pas commercialiser le nouveau médicament (i.e. dire que $\theta > \theta_0$) à tort. On rappelle que le choix de H_0 et H_1 est motivé par le fait que l'on cherche à minimiser l'erreur de 1$^{\text{ère}}$ espèce en priorité : on ne souhaite pas rejeter H_0 (et donc accepter H_1) à tort. Ceci impose donc le choix de l'hypothèse nulle $H_0 = \{\theta \leq \theta_0\}$ et de l'hypothèse alternative $H_1 = \{\theta > \theta_0\}$.

Le modèle est exponentiel et sa densité s'écrit :

$$p_n(x;\theta) = \theta^{\sum_{k=1}^n x_k}(1-\theta)^{n-\sum_{k=1}^n x_k}, \quad x = (x_1,\dots,x_n) \in \{0,1\}^n.$$

On peut considérer la statistique canonique $n\bar{X}_n$ où $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$, ce qui est cohérent avec le fait que le test porte sur la valeur du paramètre θ. Il est en effet pertinent de choisir une statistique de test construite à l'aide d'une bonne estimation du paramètre, par exemple son estimateur du maximum de vraisemblance qui ici est $\bar{X}_n$. Toutefois, pour effectuer une approche asymptotique, on considère plutôt la statistique de test :

$$\zeta_n = \sqrt{n}\frac{\bar{X}_n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}},$$

qui se déduit de $n\bar{X}_n$ par une transformation croissante. Ce choix simplifie l'étude du comportement asymptotique de la statistique de test sous H_0, comme on le voit dans ce qui suit.

On déduit de la loi forte des grands nombres que pour $\theta < \theta_0$, $\bar{X}_n - \theta_0$ converge p.s. vers $\theta - \theta_0 < 0$. On en déduit donc que pour $\theta < \theta_0$, $\mathbb{P}_\theta$-p.s. $\lim_{n\to\infty} \zeta_n = -\infty$. Pour $\theta = \theta_0$, on déduit du TCL la convergence en loi de $(\zeta_n, n \geq 1)$ vers G de loi gaussienne $\mathcal{N}(0,1)$.

De manière similaire, sous H_1 (i.e. $\theta > \theta_0$), on a $\mathbb{P}_\theta$-p.s. $\lim\limits_{n\to\infty} \zeta_n = +\infty$.

Autrement dit, pour des échantillons de grande taille, la statistique de test prend de très grandes valeurs sous H_1 et sous H_0 des valeurs très négatives ou des valeurs typiques d'une loi $\mathcal{N}(0,1)$. Comme la région critique correspond aux valeurs aberrantes sous H_0 et aux valeurs typiques sous H_1, on considère des régions critiques de la forme $W_n = \{\zeta_n \geq a\}$.

Il faut ensuite calculer le niveau asymptotique de ce test. On remarque que pour $\theta < \theta_0$ on a $\lim_{n\to+\infty} \mathbb{P}_\theta(W_n) = \lim_{n\to\infty} \mathbb{P}_\theta(\zeta_n \geq a) = 1$ et pour $\theta = \theta_0$ on a $\lim_{n\to+\infty} \mathbb{P}_{\theta_0}(W_n) = \lim_{n\to\infty} \mathbb{P}_{\theta_0}(\zeta_n \geq a) = \mathbb{P}(G \geq a)$. On en déduit que si $a = \phi_{1-\alpha}$, où $\phi_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\mathcal{N}(0,1)$, alors le test de région critique :

$$W_n = \{\zeta_n \geq \phi_{1-\alpha}\}$$

est de niveau asymptotique (non-uniforme) α. On démontre que le niveau asymptotique est en fait uniforme. Dans de nombreux exemples, on peut démontrer que l'erreur de 1ère espèce est maximale à la frontière de H_0. On remarque que sous $\mathbb{P}_\theta$, $n\bar{X}_n$ a même loi que $\sum_{k=1}^n \mathbf{1}_{\{U_k \leq \theta\}}$ où les variables aléatoires $(U_n, n \in \mathbb{N}^*)$ sont indépendantes de loi uniforme sur $[0,1]$. En particulier, pour tout $c \in \mathbb{R}$, et $\theta \leq \theta_0$, on a :

$$\begin{aligned}\mathbb{P}_\theta(n\bar{X}_n \geq c) &= \mathbb{P}(\sum_{k=1}^n \mathbf{1}_{\{U_k \leq \theta\}} \geq c) \\ &\leq \mathbb{P}(\sum_{k=1}^n \mathbf{1}_{\{U_k \leq \theta_0\}} \geq c) = \mathbb{P}_{\theta_0}(n\bar{X}_n \geq c).\end{aligned}$$

On en déduit donc que, pour $\theta \leq \theta_0$:

$$\begin{aligned}\mathbb{P}_\theta(\zeta_n \geq a) &= \mathbb{P}_\theta\left(n\bar{X}_n \geq n\theta_0 + \sqrt{n}\sqrt{\theta_0(1-\theta_0)}a\right)\\ &\leq \mathbb{P}_{\theta_0}\left(n\bar{X}_n \geq n\theta_0 + \sqrt{n}\sqrt{\theta_0(1-\theta_0)}a\right)\\ &= \mathbb{P}_{\theta_0}(\zeta_n \geq a)\end{aligned}$$

et donc $\sup_{\theta\in H_0} \mathbb{P}_\theta(W_n) = \mathbb{P}_{\theta_0}(W_n)$. L'erreur de 1ère espèce est bien maximale à la frontière de H_0. Le test est de niveau asymptotique uniforme α car :

$$\lim_{n\to+\infty} \sup_{\theta\in H_0} \mathbb{P}_\theta(W_n) = \lim_{n\to\infty} \mathbb{P}_{\theta_0}(\zeta_n \geq \phi_{1-\alpha}) = \mathbb{P}(G \geq \phi_{1-\alpha}) = \alpha.$$

On étudie ensuite le comportement du test sous H_1. Comme la statistique de test ζ_n converge p.s. vers $+\infty$ sous H_1, on en déduit que pour tout $\theta \in H_1$, $\lim_{n\to+\infty} \mathbb{P}_\theta(W_n) = \lim_{n\to+\infty} \mathbb{P}_\theta(\zeta_n \geq \phi_{1-\alpha}) = 1$. Le test est donc convergent.

Soit ζ_n^{obs} la statistique de test calculée avec les n observations (i.e. on remplace $\bar{X}_n$ dans la définition de ζ_n par $\bar{x}_n^{\text{obs}} = \frac{1}{n}\sum_{i=1}^n x_i^{\text{obs}}$, où $x_1^{\text{obs}}, \ldots, x_n^{\text{obs}}$ sont les n observations). La p-valeur asymptotique uniforme (car (IX.5) est ici vérifiée) est, d'après (IX.7) :

$$p\text{-val} = \mathbb{P}(G \geq \zeta_n^{\text{obs}}).$$

Ceci termine la construction du test. On rappelle que si la p-valeur est faible (inférieure à 1% ou de l'ordre de quelques %), on rejette H_0. ◊

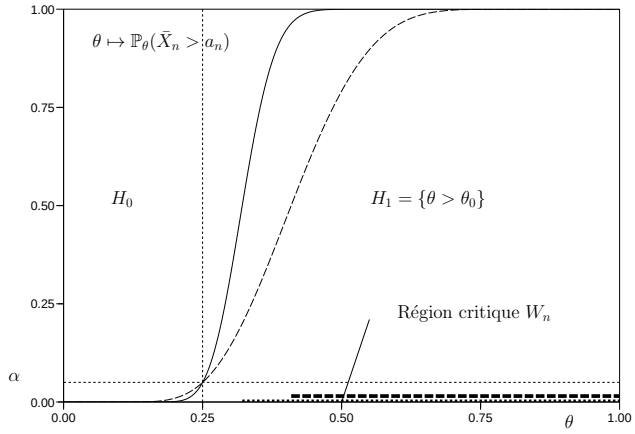


Figure IX.4. Test du modèle de Bernoulli avec les hypothèses $H_0 = \{\theta \leq \theta_0\}$ et $H_1 = \{\theta > \theta_0\}$ où $\theta_0 = 0.25$, et le niveau asymptotique $\alpha = 5\%$. La région critique du test est $W_n = \{\bar{X}_n \geq a_n\}$, avec $a_n = \theta_0 + \sqrt{\theta_0(1-\theta_0)/n}\ \phi_{1-\alpha}$, et le graphe de la fonction $\theta \mapsto \mathbb{P}_\theta(W_n)$ (erreur de 1ère sur H_0 et puissance sur H_1) sont représentés pour $n = 20$ (traits pointillés) et $n = 100$ (traits pleins).

On fait quelques remarques sur l'exemple précédent.

Remarque IX.26. La figure IX.4 représente, pour un niveau asymptotique α donné, l'erreur de 1ère espèce du test en fonction de $\theta \in H_0$ et la puissance en fonction de $\theta \in H_1$, à n fixé. On peut remarquer les comportements suivants :

- En fonction de θ. Plus θ est proche de 0 (i.e. plus on est loin de H_1), plus l'erreur de 1ère espèce est faible. De même, plus θ est proche de 1 (i.e. plus on est loin de H_0), plus la puissance est élevée et l'erreur de 2ème espèce faible. Quand $\theta \in H_1$ se rapproche de θ_0 (i.e. de H_0), on remarque que l'erreur de 2ème espèce croît vers $1-\alpha$. L'erreur de 2ème espèce est alors loin d'être négligeable. (On pourra vérifier que les tests présentés dans les différents chapitres possèdent ce comportement.)
- En fonction de la taille n de l'échantillon. La région critique, pour un niveau donné, est écrite en fonction de la statistique de test $\bar{X}_n$ au lieu de ζ_n : $W_n = \{\bar{X}_n \geq a_n\}$ avec la constante $a_n = \theta_0 + \sqrt{\theta_0(1-\theta_0)/n}\ \phi_{1-\alpha}$ qui dépend de n. Comme $\bar{X}_n$ est un estimateur convergent de θ, la région critique W_n converge quand n tend vers l'infini vers $]\theta_0, 1]$. La région critique W_n ne coïncide pas avec H_1 car il faut tenir compte *a priori* de l'erreur aléatoire de l'estimation (estimée ici par le TCL). Enfin quand n tend vers l'infini, on observe que la fonction $\theta \mapsto \mathbb{P}_\theta(W_n)$ converge vers 0 sur $[0, \theta_1[$, vers α en θ_0 et vers 1 sur $]\theta_0, 1]$.

Si le laboratoire a une démarche purement mercantile, il peut choisir de commercialiser le nouveau médicament sauf s'il est avéré que ce dernier est moins efficace que le médicament de référence. Dans ce cas on choisit $H_0 = \{\theta \geq \theta_0\}$ et $H_1 = \{\theta < \theta_0\}$. Des calculs similaires à ceux de l'exemple précédent assurent que le test asymptotique de région critique

$$W'_n = \{\zeta_n \leq \phi_\alpha\},$$

où $\phi_\alpha = -\phi_{1-\alpha}$ est le quantile d'ordre α de la loi $\mathcal{N}(0,1)$, est de niveau asymptotique uniforme α.

Si on choisit $H_0 = \{\theta \leq \theta_0\}$, on commercialise le nouveau médicament seulement si $\zeta_n \geq \phi_{1-\alpha}$; alors que si on choisit $H_0 = \{\theta \geq \theta_0\}$, on commercialise le nouveau médicament dès que $\zeta_n > -\phi_{1-\alpha}$. On voit donc que le choix des hypothèses a un impact important sur la décision finale. On remarque que si $\zeta_n^{\text{obs}} \notin]-\phi_{1-\alpha}, \phi_{1-\alpha}]$, alors la décision au niveau α ne dépend pas du choix de l'hypothèse nulle $H_0 = \{\theta \leq \theta_0\}$ ou $H_0 = \{\theta \geq \theta_0\}$. En revanche si $\zeta_n^{\text{obs}} \in]-\phi_{1-\alpha}, \phi_{1-\alpha}]$, alors la décision au niveau α dépend du choix de l'hypothèse nulle. ◊

Les étapes d'un test, qui apparaissent dans l'exemple IX.25, sont les suivantes :

1. Modélisation : le choix du modèle dépend du problème concret considéré.
2. Choix des hypothèses : il est guidé par le fait que l'on ne souhaite pas rejeter H_0 à tort (majoration *a priori* de l'erreur de 1ère espèce).
3. Choix d'une statistique de test.
4. Comportement (asymptotique ou non) de la statistique de test sous H_0.
5. Comportement (asymptotique ou non) de la statistique de test sous H_1.
6. Région critique du test : elle correspond aux valeurs aberrantes de la statistique de test sous H_0 et aux valeurs raisonnables sous H_1.
7. Contrôle de l'erreur de 1ère espèce : détermination à partir du point 4 de la région critique en fonction du niveau du test (valeur exacte, majoration, valeur asymptotique ou valeur asymptotique uniforme).
8. Contrôle de l'erreur de 2ème espèce : analyse à partir du point 5 de la puissance ou convergence (approche asymptotique) du test.
9. Détermination de la p-valeur (valeur exacte, majoration, valeur asymptotique ou valeur asymptotique uniforme).

La p-valeur permet de conclure : on rejette H_0 si elle est "**faible**". La notion "faible" dépend du contexte ; en général "faible" signifie inférieur à quelques %.

Exemple. Test de la moyenne pour un échantillon gaussien à variance connue. On considère un échantillon gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$ à variance connue. On souhaite tester $H_0 = \{\mu_0\}$ contre $H_1 = \{\mu \neq \mu_0\}$. On considère la statistique de test $\zeta_n = \sqrt{n}(\bar{X}_n - \mu_0)/\sigma_0$ où $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Sous H_0, ζ_n est de loi $\mathcal{N}(0,1)$ et sous H_1, ζ_n est de loi gaussienne de moyenne non nulle. On choisit donc une région critique de la forme $W_n = \{|\zeta_n| \geq a\}$. Ce test est de niveau exact α pour $a = \phi_{1-\alpha/2}$, le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. Ce test correspond au test UPPS de l'exemple IX.19. Par la loi forte des grands nombres, $\bar{X}_n$ converge $\mathbb{P}_{\mu_1}$-p.s. vers μ_1. Donc si $\mu_1 \neq \mu_0$, on a $\mathbb{P}_{\mu_1}$-p.s. $\lim_{n\to\infty} \zeta_n \in \{-\infty, +\infty\}$. On déduit du théorème de convergence dominée que le test est convergent. La p-valeur du test est donnée par $\mathbb{P}(|G| \geq |\zeta_n^{\text{obs}}|)$, où G est de loi $\mathcal{N}(0,1)$ et ζ_n^{obs} est la statistique de test évaluée sur les observations. D'après la proposition III.22, la p-valeur est sous H_0 la réalisation d'une variable aléatoire de loi uniforme sur $[0,1]$. ◊

IX.8.2 Hypothèse implicite : le test de Wald

Le test de Wald est défini dans un modèle paramétrique pour l'hypothèse nulle implicite : $H_0 = \{g(\theta) = 0\}$ où g est une fonction du paramètre θ.

1. Le modèle.

On considère un modèle d'échantillonnage pour un modèle paramétrique $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ régulier et identifiable. On suppose de plus que l'information de Fisher est une fonction continue du paramètre, inversible et d'inverse continue. Ceci est vérifié dans de nombreux exemples.

2. Les hypothèses.

On suppose que l'ensemble des paramètres Θ est un ouvert de $\mathbb{R}^p$. Soit g une fonction définie sur Θ à valeurs réelles ou vectorielles. On souhaite tester l'hypothèse nulle **implicite** $H_0 = \{\theta \in \Theta; g(\theta) = 0\}$ contre l'hypothèse alternative $H_1 = \{\theta \in \Theta; g(\theta) \neq 0\}$.

3. La statistique de test.

Pour tester une hypothèse implicite, il existe plusieurs tests asymptotiques (test du rapport de vraisemblance, test du score, test de Hausman, test du multiplicateur de Lagrange, ...). On considère dans ce qui suit le **test de Wald**.

On suppose que la fonction g est une fonction de classe C^1 de $\mathbb{R}^p$ dans $\mathbb{R}^r$ et que la matrice $\dfrac{\partial g}{\partial \theta}(\theta) = \left(\dfrac{\partial g_i}{\partial \theta_j}(\theta); 1 \leq i \leq r, 1 \leq j \leq p\right)$ de taille $r \times p$ est de **rang** r pour tout $\theta \in \Theta$.

Remarque. Si $p = r = 1$, et si Θ est connexe, alors la fonction g est soit strictement croissante soit strictement décroissante. Donc il existe au plus une racine θ_0 à l'équation $g(\theta) = 0$. Si elle existe alors $H_0 = \{\theta_0\}$ et $H_1 = \Theta \backslash \{\theta_0\}$. Si on suppose que l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est convergent, alors sous H_0 il converge p.s. vers θ_0 et par continuité, $g(\hat{\theta}_n)$ converge p.s. vers $g(\theta_0) = 0$. $\Diamond$

Intuitivement l'ensemble des racines de $g(\theta) = 0$ forme un "espace" (en fait une sous-variété) de dimension $p - r$. Le test de Wald consiste à regarder l'écart entre le vecteur $g(\hat{\theta}_n)$ et 0, où $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance du modèle d'échantillonnage de taille n. On considère la statistique de test :

$$\zeta_n = n\, g(\hat{\theta}_n)^t \Sigma(\hat{\theta}_n)^{-1} g(\hat{\theta}_n) \quad \text{avec } \Sigma(\theta) = \frac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1} \left(\frac{\partial g}{\partial \theta}(\theta)\right)^t, \tag{IX.9}$$

où $I(\theta)$ est l'information de Fisher du modèle.

Remarque IX.27. On ne dispose pas toujours de la forme explicite de $I(\theta)$. On peut alors estimer $I(\theta)$ en utilisant la méthode des moments et l'une des deux formules :

$$I(\theta) = \mathbb{E}_\theta \left[\frac{\partial \log p(X_1;\theta)}{\partial \theta} \left(\frac{\partial \log p(X_1;\theta)}{\partial \theta} \right)^t \right] \text{ ou } \quad I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log p(X_1;\theta)}{\partial \theta^2} \right].$$

On peut donc remplacer la matrice $I(\cdot)$ dans (IX.9) par une des deux estimations convergentes sans biais suivantes :

$$\hat{I}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log p(X_i;\cdot)}{\partial \theta} \left(\frac{\partial \log p(X_i;\cdot)}{\partial \theta} \right)^t \text{ ou } \quad \hat{I}(\cdot) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log p(X_i;\cdot)}{\partial \theta^2}.$$

$\Diamond$

4-5. Comportement sous H_0 et H_1.

Le comportement de la statistique de test est donné par la proposition suivante.

Proposition IX.28 (Wald)**.** *Soit $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. La statistique de test ζ_n, définie par (IX.9) converge en loi* ***sous*** *H_0 vers un χ^2 à r* ***degrés de liberté*** *et elle converge en probabilité* ***sous*** *H_1 vers $+\infty$.*

Démonstration. On rappelle que l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est un estimateur asymptotiquement efficace de θ d'après le théorème VIII.42. En particulier sa matrice de covariance asymptotique est $I(\theta)^{-1}$. Grâce à la proposition VIII.40, on en déduit que $g(\hat{\theta}_n)$ est un estimateur de $g(\theta)$ asymptotiquement normal de matrice de covariance $\Sigma(\theta)$. Donc $\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right)$ converge en loi vers un vecteur gaussien G de loi $\mathcal{N}(0, \Sigma(\theta))$. Sous H_0, comme $g(\theta) = 0$, on a la convergence en loi de $\sqrt{n}\, g(\hat{\theta}_n)$ vers G. On remarque que la matrice $\Sigma(\theta)$ est régulière car la matrice $\dfrac{\partial g}{\partial \theta}$ est de rang r. De plus l'application $\theta \mapsto \Sigma(\theta)$ est continue. On en déduit que l'application $\theta \mapsto \Sigma(\theta)^{-1/2}$ est continue. Le théorème de Slutsky (pour des variables aléatoires vectorielles) implique que sous H_0, $\Sigma(\hat{\theta}_n)^{-1/2}\sqrt{n}\, g(\hat{\theta}_n)$ converge en loi vers $Y = \Sigma(\theta)^{-1/2} G$ de loi $\mathcal{N}(0, I_r)$ où I_r est la matrice identité de taille $r \times r$. Par conséquent,

$$\zeta_n = n\, g(\hat{\theta}_n)^t \Sigma(\hat{\theta}_n)^{-1} g(\hat{\theta}_n) = \left| \Sigma(\hat{\theta}_n)^{-1/2} \sqrt{n}\, g(\hat{\theta}_n) \right|^2$$

converge en loi vers $|Y|^2$ c'est-à-dire un χ^2 à r degrés de liberté.

Le théorème VIII.42 et la remarque VIII.39 impliquent que $g(\hat{\theta}_n)$ converge en probabilité vers $g(\theta)$. On en déduit que sous H_1, $\frac{1}{n}\zeta_n$ converge en probabilité vers $g(\theta)^t \Sigma(\theta)^{-1} g(\theta) \neq 0$. Donc sous H_1, la quantité ζ_n diverge en probabilité. □

6. La région critique.

La statistique de test prend de grandes valeurs sous H_1 et sous H_0 des valeurs typiques d'une loi $\chi^2(r)$. On choisit donc des régions critiques de la forme $W_n = \{\zeta_n \geq a\}$.

7. Contrôle de l'erreur de 1ère espèce.

On déduit de la convergence en loi sous H_0 de ζ_n dans la proposition IX.28 que le test de région critique $W_n = \{\zeta_n \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(r)$ est de niveau asymptotique (non-uniforme) α.

8. Contrôle de l'erreur de 2ème espèce.

On déduit de la convergence en probabilité vers $+\infty$ de ζ_n sous H_1 et de la proposition V.19 que le test est convergent.

9. La p-valeur.

Soit ζ_n^{obs} la statistique de test calculée avec les n observations. Soit Z de loi $\chi^2(r)$. On déduit de (IX.8) que la p-valeur asymptotique (non-uniforme car seulement (IX.6) est vérifiée) est :

$$p\text{-val} = \sup_{\theta \in H_0} \lim_{n\to\infty} \mathbb{P}_\theta(\zeta_n \geq \zeta_n^{\text{obs}}) = \mathbb{P}(Z \geq \zeta_n^{\text{obs}}).$$

Les résultats de 4-9 sont conservés si on remplace I par une des approximations de la remarque IX.27.

L'exemple et les exercices qui suivent concernent des tests d'égalité de moyenne pour deux échantillons gaussiens indépendants.

Exemple IX.29. Soit un échantillon gaussien (X_i, Y_i) de loi $\mathcal{N}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix}\right)$, où σ_1 et σ_2 sont connus. Le paramètre est $\theta = (\mu_1, \mu_2)$. On souhaite tester l'égalité $\mu_1 = \mu_2$ i.e. l'hypothèse implicite $H_0 = \{\mu_1 - \mu_2 = 0\}$ contre son alternative

$H_1 = \{\mu_1 \neq \mu_2\}$. On pose $g(\mu_1, \mu_2) = \mu_1 - \mu_2$. On remarque qu'il s'agit d'un modèle exponentiel sous sa forme naturelle. On obtient en utilisant le test de Wald et les notations $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^n Y_i$, que $\zeta_n = \frac{n(\hat{\mu}_1 - \hat{\mu}_2)^2}{\sigma_1^2 + \sigma_2^2}$ et que, sous H_0, ζ_n converge en loi vers un χ^2 à 1 degré de liberté. On vérifie que dans ce cadre, on a en fait $\mathcal{L}(\zeta_n) = \chi^2(1)$. Un test convergent de niveau exact (et bien sûr asymptotique) 5% est donné par la région critique $W_n = \{\zeta_n \geq 3,84\}$, où 3,84 est le quantile d'ordre 95% de la loi $\chi^2(1)$ (voir la table du χ^2, paragraphe XI.3). La p-valeur exacte de ce test est $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$, où Z est de loi $\chi^2(1)$ et ζ_n^{obs} la statistique de test évaluée en les données. Sous H_0, la p-valeur est la réalisation d'une v.a. de loi uniforme sur $[0, 1]$. ◊

Exercice IX.2.
Reprendre l'exemple IX.29 en supposant que σ_1 et σ_2 sont inconnus.

1. Vérifier que le test de Wald donne la même région critique en remplaçant σ_1^2 et σ_2^2 par leur estimateur du maximum de vraisemblance : $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$, mais que le niveau de ce test est asymptotique (non-uniforme).
2. Montrer que si $\sigma_1 = \sigma_2 = \sigma$ (avec σ inconnu), alors ζ_n a même loi que $2n\frac{Z_1}{Z_{2n-2}}$, où les variables Z_1 et Z_{2n-2} sont indépendantes de lois respectives $\chi^2(1)$ et $\chi^2(2n-2)$ et donc que $\frac{n-1}{n}\zeta_n$ suit une loi de Fisher-Snedecor de paramètre $(1, 2n-2)$. Donner alors une région critique de niveau exact pour tester $H_0 = \{\mu_1 = \mu_2\}$ contre son alternative $H_1 = \{\mu_1 \neq \mu_2\}$. △

Exercice IX.3.
On dispose de deux échantillons de tailles différentes, indépendants $(X_i; 1 \leq i \leq n_1)$ et $(Y_j; 1 \leq j \leq n_2)$ de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$ où μ_i et σ_i sont inconnus. Déduire de l'exercice IX.2 un test asymptotique lorsque $\min(n_1, n_2)$ tend vers $+\infty$, pour l'hypothèse nulle $H_0 = \{\mu_1 = \mu_2\}$ contre son alternative $H_1 = \{\mu_1 \neq \mu_2\}$ construit à l'aide de la statistique de test :

$$\zeta_{n_1,n_2} = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}.$$

△

IX.8.3 Hypothèse explicite : le test de Hausman

Le test de Hausman est défini dans un modèle paramétrique pour l'hypothèse nulle explicite : $H_0 = \{\theta = h(\gamma)\}$ où le paramètre θ peut se réécrire à l'aide d'une fonction h d'un autre paramètre γ de dimension plus petite.

1. Le modèle.

On considère un modèle d'échantillonnage pour un modèle paramétrique $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$ régulier et identifiable. On suppose de plus que l'information de Fisher est une fonction continue du paramètre, inversible et d'inverse continue. Ceci est vérifié dans de nombreux exemples.

2. Les hypothèses.

On suppose que l'ensemble des paramètres Θ est un ouvert de $\mathbb{R}^p$. Soit h une fonction définie sur un ouvert Γ de $\mathbb{R}^q$ à valeurs dans Θ. On souhaite tester l'hypothèse nulle **explicite** $H_0 = \{\theta = h(\gamma); \gamma \in \Gamma\}$ contre l'hypothèse alternative $H_1 = \{\theta \notin h(\Gamma)\}$.

3. Les statistiques de test.

On suppose que la fonction h est injective, de classe C^1 et que la matrice $\dfrac{\partial h}{\partial \gamma}(\gamma) = \left(\dfrac{\partial h_j}{\partial \gamma_i}(\gamma); 1 \leq i \leq q, 1 \leq j \leq p\right)$ est de rang $q < p$ pour tout $\gamma \in \Gamma$.

Le **test de Hausman** est construit à partir de la différence entre $\hat{\theta}_n$ et $h(\hat{\gamma}_n)$, où $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ pour un échantillon de taille n et $\hat{\gamma}_n$ l'estimateur du maximum de vraisemblance de γ pour un échantillon de taille n. Pour l'estimation de γ on considère le modèle paramétrique $\mathcal{P}' = \{\mathrm{P}_{h(\gamma)}; \gamma \in \Gamma\}$. On remarque que $h(\hat{\gamma}_n)$ est l'estimateur du maximum de vraisemblance de θ sous la contrainte $\theta \in h(\Gamma) \subset \Theta$. Grâce aux propriétés des estimateurs du maximum de vraisemblance, on a que sous H_0, si $\theta_0 = h(\gamma_0)$ est le vrai paramètre de la loi de l'échantillon, alors $\hat{\theta}_n - h(\hat{\gamma}_n)$ converge p.s. vers $\theta_0 - h(\gamma_0) = 0$. Les deux statistiques de test suivantes considèrent cette différence renormalisée :

$$\zeta_n^{(1)} = n(\hat{\theta}_n - h(\hat{\gamma}_n))^t I(\hat{\theta}_n)(\hat{\theta}_n - h(\hat{\gamma}_n)), \tag{IX.10}$$

$$\zeta_n^{(2)} = n(\hat{\theta}_n - h(\hat{\gamma}_n))^t I(h(\hat{\gamma}_n))(\hat{\theta}_n - h(\hat{\gamma}_n)), \tag{IX.11}$$

où $I(\theta)$ est l'information de Fisher du modèle. Voir la remarque IX.27 pour l'estimation de la fonction I, quand on ne peut pas la calculer directement.

4-5. Comportement sous H_0 et H_1.

Le comportement des statistiques de test est donné par la proposition suivante que l'on admet.

Proposition IX.30 (Hausman)**.** *Soit $\hat{\theta}_n$ et $\hat{\gamma}_n$ les estimateurs du maximum de vraisemblance de θ et γ. On considère les statistiques de test définies par (IX.10) et (IX.11). Le couple $(\zeta_n^{(1)}, \zeta_n^{(2)})$ converge en loi* ***sous*** *H_0 vers (Z, Z) où Z est de loi χ^2 à $p-q$* ***degrés de liberté*** *et les statistiques de test convergent en probabilité* ***sous*** *H_1 vers $+\infty$.*

Le nombre de degrés de liberté du test du χ^2 est égal à la dimension du paramètre moins le nombre d'estimations pour la contrainte γ.

6. La région critique.

Les statistiques de test prennent de très grandes valeurs sous H_1 et sous H_0 des valeurs typiques d'une loi $\chi^2(p-q)$. On considère donc des régions critiques de la forme $W_n^{(i)} = \left\{\zeta_n^{(i)} \geq a\right\}$ pour $i \in \{1, 2\}$.

Un raisonnement similaire à celui du paragraphe IX.8.2 permet d'obtenir les résultats suivants.

7. Contrôle de l'erreur de 1ère espèce.

Pour $i \in \{1, 2\}$, le test de région critique $W_n^{(i)} = \left\{\zeta_n^{(i)} \geq z_{1-\alpha}\right\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(p-q)$ est de niveau asymptotique (non-uniforme) α.

8. Contrôle de l'erreur de 2ème espèce.

Les tests sont convergents.

9. La p-valeur.

Soit $\zeta_n^{(i),\text{obs}}$ la statistique de test $\zeta_n^{(i)}$ calculée avec les n observations. Les p-valeurs non-uniformes asymptotiques sont, pour $i \in \{1, 2\}$:

$$p\text{-val}^{(i)} = \sup_{\theta \in H_0} \lim_{n \to \infty} \mathbb{P}_\theta(\zeta_n^{(i)} \geq \zeta_n^{(i),\text{obs}}) = \mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}}),$$

où Z est de loi $\chi^2(p-q)$.

Les résultats de 4-9 sont conservés si on remplace I par une des approximations de la remarque IX.27.

Exercice IX.4.
Reprendre l'exemple IX.29 en utilisant le test de Hausman. Vérifier que l'on a $\zeta_n^{(1)} = \zeta_n^{(2)} = n\dfrac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{\sigma_1^2 + \sigma_2^2}$. Dans ce cas le test de Hausman est égal au test de Wald. △

IX.9 Test d'adéquation du χ^2 et applications

Le test de Hausman est particulièrement bien adapté pour vérifier l'adéquation des données à une famille de lois discrètes à support fini. Après avoir donné un résultat général au paragraphe IX.9.1, on considère les cas particuliers importants en pratique : le test d'adéquation à une loi au paragraphe IX.9.2 et le test d'indépendance au paragraphe IX.9.3. Le paragraphe IX.9.4 est consacré à la démonstration de la proposition IX.31.

IX.9.1 Test du χ^2 empirique

Le test du χ^2 empirique est défini pour des variables aléatoires prenant un nombre fini m de valeurs et l'hypothèse nulle correspond à une famille de lois qui dépendent d'un paramètre de dimension q inférieure ou égale à $m-1$.

1. Le modèle.

On considère le modèle d'échantillonnage suivant : les variables aléatoires $(X_n, n \in \mathbb{N}^*)$ sont indépendantes de même loi et à valeurs dans $\{a_1, \ldots, a_m\}$, où $a_1, \ldots, a_m$ sont distincts. On note P_p la loi de X_1 qui dépend du paramètre $p = (p_1, \ldots, p_m)$ où p_i est la probabilité que X_1 soit égal à a_i. Plus précisément, on considère le modèle paramétrique $\mathcal{P} = \{\mathrm{P}_p; p \in Q\}$ avec $Q = \{p = (p_1, \ldots, p_m); p_i > 0, \text{ pour } 1 \leq i \leq m \quad \text{et} \quad \sum_{i=1}^m p_i = 1\}$. Par hypothèse la probabilité $p_i = \mathbb{P}_p(X_1 = i)$ est strictement positive pour tout $i \in \{1, \ldots, m\}$.

On remarque que le paramètre p appartient à un sous-espace affine de dimension $m-1$ à cause de la **contrainte** $\sum_{i=1}^m p_i = 1$.

2. Les hypothèses.

On se donne une partie de Q caractérisée par $p = p(\gamma) = h(\gamma)$ où la fonction h est définie sur un **ouvert** $\Gamma \subset \mathbb{R}^q$ et à valeurs dans Q, avec $0 \leq q < m-1$.

On considère l'hypothèse nulle $H_0 = \{p = h(\gamma); \gamma \in \Gamma\}$ et l'hypothèse alternative $H_1 = Q \backslash H_0$.

3. Les statistiques de test.

On suppose que h est injective de classe C^1 et que la matrice $\dfrac{\partial h}{\partial \gamma}(\gamma)$ est de rang q pour tout $\gamma \in \Gamma$. On considère les statistiques de test sur l'échantillon $X_1, \ldots, X_n$ de taille n :

$$\boxed{\zeta_n^{(1)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{\hat{p}_i}} \quad \text{et} \quad \boxed{\zeta_n^{(2)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{p_i(\hat{\gamma}_n)},} \tag{IX.12}$$

où $\hat{p}$ est le vecteur des fréquences empiriques : $\boxed{\hat{p} = \left(\frac{N_1}{n}, \ldots, \frac{N_m}{n}\right)}$ avec les occurrences $N_i = \sum_{k=1}^{n} \mathbf{1}_{\{X_k = i\}}$ et où $\hat{\gamma}_n$ est l'estimateur du maximum de vraisemblance de γ sur l'échantillon de taille n. Le corollaire IX.32 au paragraphe IX.9.4 assure que le vecteur des fréquences empiriques $\hat{p}$ est l'estimateur du maximum de vraisemblance du vecteur des fréquences p.

4-5. Comportement sous H_0 et H_1.

La proposition IX.30 permet d'obtenir le comportement des statistiques de test. Sa démonstration est reportée au paragraphe IX.9.4.

Proposition IX.31 (Test du χ^2 empirique)**.** *On considère les statistiques de test définies par (IX.12). Le couple $(\zeta_n^{(1)}, \zeta_n^{(2)})$ converge en loi* ***sous*** *H_0 vers (Z, Z) où Z est de loi χ^2 à $m-1-q$* ***degrés de liberté*** *et les statistiques de test convergent en probabilité* ***sous*** *H_1 vers $+\infty$.*

6. La région critique.

Les statistiques de test prennent de très grandes valeurs sous H_1 et sous H_0 des valeurs typiques d'une loi $\chi^2(m-1-q)$. On considère donc des régions critiques de la forme $W_n^{(i)} = \left\{\zeta_n^{(i)} \geq a\right\}$ pour $i \in \{1, 2\}$.

Un raisonnement similaire à celui du paragraphe IX.8.2 permet d'obtenir les résultats suivants.

7. Contrôle de l'erreur de 1ère espèce.

Pour $i \in \{1,2\}$, le test de région critique $W_n^{(i)} = \left\{\zeta_n^{(i)} \geq z_{1-\alpha}\right\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(m-1-q)$ est de niveau asymptotique (non-uniforme) α.

8. Contrôle de l'erreur de 2ème espèce.

Les tests sont convergents.

9. La p-valeur.

Soit $\zeta_n^{(i),\text{obs}}$ la statistique de test $\zeta_n^{(i)}$ calculée avec les n observations. Les p-valeurs non-uniformes asymptotiques sont, pour $i \in \{1,2\}$:

$$p\text{-val}^{(i)} = \sup_{\theta \in H_0} \lim_{n\to\infty} \mathbb{P}_\theta(\zeta_n^{(i)} \geq \zeta_n^{(i),\text{obs}}) = \mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}}),$$

où Z est de loi $\chi^2(m-1-q)$.

Remarque.

- Le test du χ^2 est un test asymptotique. On regarde donc des échantillons dont la taille tend vers l'infini. Dans la pratique les échantillons sont de tailles limitées. La question se pose de la validité du test du χ^2. Une attitude conservatrice consiste à utiliser l'approximation asymptotique dès que $\inf_i N_i \geq 5$ voire 10 ou $\inf_i np_i \geq 5$ voire 10. On désire donc avoir 5 ou 10 éléments dans chaque groupe, soit sur la réalisation soit en théorie. Si cela n'est pas le cas, alors on peut regrouper des groupes de manière à satisfaire le critère ci-dessus. Ce faisant, le test devient moins puissant. Le mieux étant certainement de faire des simulations pour avoir des estimations du niveau du test.
- Pour se rappeler du nombre de degrés de liberté, on compte le nombre de paramètres (ici on a donc m paramètres), auquel on retranche le nombre de contraintes et le nombre de paramètres à estimer pour γ. Ici la contrainte est $\sum_{i=1}^m p_i = 1$ et comme $\gamma \in \mathbb{R}^q$, il faut estimer $\gamma_1, \ldots, \gamma_q$: il reste donc $m-1-q$ degrés de liberté.

$\Diamond$

Exercice IX.5.
Le tableau IX.4 donne[1], sur une période de vingt ans (1875-1894), le nombre de

1. A. Gulberg, Les fonctions de fréquence discontinues et les séries statistiques, *Annales de l'Inst. H. Poincaré.*, 3, pp.229-278, 1933.

décès par an et par régiment dans la cavalerie prussienne causés par un coup de sabot de cheval. On dispose de 280 observations. Appliquer le test du χ^2 pour vérifier si les données suivent une loi de Poisson (dont on estimera le paramètre). △

Nombre de décès par an et par régiment	0	1	2	3	4
Nombre d'observations	144	91	32	11	2

Table IX.4. Décès par an et par régiment.

IX.9.2 Test d'adéquation à une loi

Le test d'adéquation de loi est un test du χ^2 empirique où l'hypothèse nulle est réduite à une seule loi d'une variable aléatoire prenant un nombre fini m de valeurs.

1. On considère le modèle et les notations du paragraphe IX.9.1.

2. On désire tester si la loi des variables aléatoires indépendantes $(X_n, n \in \mathbb{N}^*)$ est égale à une loi fixée P_{p^0}, où $p^0 = (p_1^0, \ldots, p_m^0) \in Q$ (i.e. $p_i^0 > 0$ pour $1 \leq i \leq m$ et $\sum_{i=1}^m p_i^0 = 1$). L'hypothèse nulle est $H_0 = \{p = p^0\}$ et l'hypothèse alternative $H_1 = \{p \in Q, p \neq p^0\}$.

3. On applique le test du χ^2 empirique pour l'échantillon de taille n avec $q = 0$ et $h(p) = p^0$. Les statistiques du χ^2 empiriques définies par (IX.12) s'écrivent :

$$\zeta_n^{(1)} = n \sum_{i=1}^m \frac{\left(\hat{p}_i - p_i^0\right)^2}{\hat{p}_i} \quad \text{et} \quad \zeta_n^{(2)} = n \sum_{i=1}^m \frac{\left(\hat{p}_i - p_i^0\right)^2}{p_i^0}.$$

4-8. On déduit du paragraphe IX.9.1 que, pour $i \in \{1, 2\}$, on rejette l'hypothèse nulle H_0 au niveau asymptotique α si $\zeta_n^{(i)}$ est supérieur au quantile d'ordre $1 - \alpha$ du χ^2 à $m - 1$ degrés de liberté et que ce test est convergent.

9. La p-valeur asymptotique non-uniforme de ce test est $p\text{-val}^{(i)} = \mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée avec les n observations et Z est de loi $\chi^2(m-1)$.

Exemple. On considère les naissances aux U.S.A. On souhaite savoir si les naissances sont uniformes sur les jours de la semaine. Les nombres moyens de naissances par jour de semaine pour 1997 sont les suivants (source "National Vital

Statistics Report" 1999) : $N_{\text{lundi}} = 10861$, $N_{\text{mardi}} = 12104$, $N_{\text{mercredi}} = 11723$, $N_{\text{jeudi}} = 11631$, $N_{\text{vendredi}} = 11640$, $N_{\text{samedi}} = 8670$ et $N_{\text{dimanche}} = 7778$.

On obtient $\zeta_n^{(1)} \simeq 1949$ et $\zeta_n^{(2)} \simeq 1639$. On lit dans la table du χ^2 que le quantile d'ordre 0.1% du $\chi^2(6)$ est 23. On rejette donc l'hypothèse des naissances uniformes sur les jours de la semaine au niveau asymptotique 0,1%. En fait les p-valeurs sont presque nulles (inférieures à 10^{-1000}), ceci justifie que l'on rejette H_0.

On remarque que pour répondre à la question et appliquer le test de χ^2, il aurait fallu utiliser le nombre de naissances le jour j en 1997 soit environ 52 N_j et non le nombre moyen N_j. Le résultat du test est alors encore plus net. ◊

IX.9.3 Test d'indépendance

Le test d'indépendance est un test du χ^2 empirique où l'hypothèse nulle correspond à la loi d'un couple (V, W) avec V et W indépendants.

1. Soit $(X_n = (V_n, W_n), n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes à valeurs dans $\{b_1, \ldots, b_\ell\} \times \{c_1, \ldots, c_r\}$. On note $I = \{1, \ldots, \ell\}$, $J = \{1, \ldots, r\}$ et, pour $(i, j) \in I \times J$, les probabilités $p_{i,j} = \mathbb{P}_p(V_1 = b_i, W_1 = c_j) > 0$ et les probabilités marginales :

$$p_{i,\cdot} = \sum_{j=1}^{r} p_{i,j} = \mathbb{P}_p(V_1 = b_i) \quad \text{et} \quad p_{\cdot,j} = \sum_{i=1}^{\ell} p_{i,j} = \mathbb{P}_p(W_1 = c_j).$$

2. On souhaite vérifier si les variables V_1 et W_1 sont indépendantes, c'est-à-dire si les probabilités $p_{i,j}$ sont les produits des probabilités marginales $p_{i,\cdot}$ et $p_{\cdot,j}$. Pour cela on utilise le test du χ^2 empirique du paragraphe IX.9.1 avec l'hypothèse nulle :

$$H_0 = \{p \in Q; p_{i,j} = p_{i,\cdot} p_{\cdot,j}, \text{ pour tout } (i,j) \in I \times J\},$$

où $Q = \{p = (p_{i,j}, (i,j) \in I \times J); p_{i,j} > 0 \text{ pour tout } (i,j) \in I \times J, \sum_{i \in I, j \in J} p_{i,j} = 1\}$ et l'hypothèse alternative $H_1 = Q \backslash H_0$.

Pour tenir compte des contraintes $\sum_{i=1}^{\ell} p_{i,\cdot} = 1$ et $\sum_{j=1}^{r} p_{\cdot,j} = 1$ sur les probabilités marginales, on considère l'ouvert de $\mathbb{R}^{(\ell-1)+(r-1)}$:

$$\Gamma = \Big\{\gamma = (\gamma_1, \ldots, \gamma_{\ell+r-2}); \\ \gamma_k > 0 \text{ pour } k \in \{1, \ldots, \ell + r - 2\}, \sum_{k=1}^{\ell-1} \gamma_k < 1, \quad \sum_{k=\ell}^{\ell-1+r-1} \gamma_k < 1\Big\}$$

et la fonction h définie sur Γ par $p_{i,j} = h_{i,j}(\gamma) = p_{i,\cdot}p_{\cdot,j}$ où $\gamma = (\gamma_1, \ldots, \gamma_{\ell+r-2})$ avec $p_{i,\cdot} = \gamma_i$ si $i \leq \ell - 1$, $p_{\ell,\cdot} = 1 - \sum_{i=1}^{\ell-1} \gamma_i$, $p_{\cdot,j} = \gamma_{\ell-1+j}$ si $j \leq r-1$ et $p_{\cdot,r} = 1 - \sum_{i=\ell}^{r+\ell-2} \gamma_i$. On remarque que les coordonnées de γ sont des fréquences.

3. La fonction h est une injection de $\Gamma \subset \mathbb{R}^{\ell+r-2}$ dans $\mathbb{R}^{r\ell}$ de classe C^1. Il est facile de vérifier que la matrice des dérivées de h est de rang la dimension de l'espace de départ : $q = \ell + r - 2$. Le nombre de degrés de liberté est donc :

$$m - 1 - q = \ell r - 1 - (\ell + r - 2) = (\ell - 1)(r - 1).$$

On note les occurrences :

$$N_{i,j} = \sum_{k=1}^{n} \mathbf{1}_{\{V_k = b_i, W_k = c_j\}}, \quad N_{i,\cdot} = \sum_{k=1}^{n} \mathbf{1}_{\{V_k = b_i\}} \quad \text{et} \quad N_{\cdot,j} = \sum_{k=1}^{n} \mathbf{1}_{\{W_k = c_j\}}.$$

La démonstration du corollaire IX.32 au paragraphe IX.9.4 assure que l'estimateur du maximum de vraisemblance de γ (dans le modèle sous H_0) est l'estimateur des fréquences empiriques :

$$\hat{\gamma}_n = \left(\frac{N_{1,\cdot}}{n}, \ldots, \frac{N_{\ell-1,\cdot}}{n}, \frac{N_{\cdot,1}}{n}, \ldots, \frac{N_{\cdot,r-1}}{n} \right).$$

On en déduit donc que $p_{i,j}(\hat{\gamma}_n) = \dfrac{N_{i,\cdot}}{n} \dfrac{N_{\cdot,j}}{n}$ pour tout $(i,j) \in I \times J$. Les statistiques du χ^2 empiriques définies par (IX.12) s'écrivent :

$$\zeta_n^{(1)} = \sum_{i=1}^{\ell} \sum_{j=1}^{r} \frac{\left(N_{i,j} - \frac{N_{i,\cdot} N_{\cdot,j}}{n} \right)^2}{N_{i,j}} \quad \text{et} \quad \zeta_n^{(2)} = n \sum_{i=1}^{\ell} \sum_{j=1}^{r} \frac{\left(N_{i,j} - \frac{N_{i,\cdot} N_{\cdot,j}}{n} \right)^2}{N_{i,\cdot} N_{\cdot,j}}.$$

4-8. On déduit du paragraphe IX.9.1, pour $i \in \{1, 2\}$, que l'on rejette l'hypothèse nulle d'indépendance H_0 au niveau asymptotique α si $\zeta_n^{(i)}$ est supérieur au quantile d'ordre $1 - \alpha$ du χ^2 à $(\ell - 1)(r - 1)$ degrés de liberté et que ce test est convergent.

9. La p-valeur asymptotique non-uniforme de ce test est $p\text{-val}^{(i)} = \mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée avec les n observations et Z est de loi $\chi^2((\ell - 1)(r - 1))$.

Exercice IX.6.
On désire étudier la répartition des naissances suivant le type du jour de semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données du tableau IX.5 proviennent du "National Vital Statistics Report" et concernent les naissances aux USA en 1997.

À l'aide d'un test du χ^2, pouvez-vous accepter ou rejeter l'hypothèse d'indépendance entre le type du jour de naissance (jours ouvrables ou week-end) et le mode d'accouchement (naturel ou césarienne) ?

Naissances	Naturelles	César.	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Table IX.5. Répartition des naissances aux USA en 1997 suivant le mode d'accouchement. (On a omis 35 240 naissances pour lesquelles le mode d'accouchement n'a pas été retranscrit.).

△

IX.9.4 Test du χ^2 empirique (démonstration)

On donne la démonstration de la proposition IX.31. On explicite pour ce modèle d'échantillonnage particulier le test de Hausman. Les résultats énoncés au paragraphe IX.8.3 requièrent que le paramètre à estimer varie dans un ouvert, or ici Q n'est pas un ouvert. Pour résoudre cette difficulté, on pose $\theta = (p_1, \dots, p_{m-1})$ et

$$\Theta = \Big\{\theta = (\theta_1, \dots, \theta_{m-1}) \in \mathbb{R}^{m-1}; \theta_i > 0 \text{ pour } 1 \leq i \leq m-1 \text{ et } \sum_{i=1}^{m-1} \theta_i < 1\Big\}.$$

Le paramètre θ est de dimension $m-1$ et Θ est un ouvert de $\mathbb{R}^{m-1}$.

Pour déterminer l'information de Fisher, on remarque que la vraisemblance et la log-vraisemblance s'écrivent :

$$p(x_1;\theta) = \theta_1^{\mathbf{1}_{\{x_1=a_1\}}} \cdots \theta_{m-1}^{\mathbf{1}_{\{x_1=a_{m-1}\}}} (1-\theta_1 \cdots - \theta_{m-1})^{\mathbf{1}_{\{x_1=a_m\}}},$$
$$L_1(x_1;\theta) = \sum_{i=1}^{m-1} \mathbf{1}_{\{x_1=a_i\}} \log(\theta_i) + \mathbf{1}_{\{x_1=a_m\}} \log\left(1 - \sum_{l=1}^{m-1} \theta_l\right).$$

On en déduit le vecteur du score $V_1 = (V_{1,1}, \dots, V_{1,m-1})$:

$$V_{1,i} = \frac{\partial L_1(X_1;\theta)}{\partial \theta_i} = \frac{1}{\theta_i} \mathbf{1}_{\{X_1=a_i\}} - \frac{1}{1 - \sum_{l=1}^{m-1} \theta_l} \mathbf{1}_{\{X_1=a_m\}},$$

et l'information de Fisher $I(\theta) = (I(\theta)_{i,j}; 1 \leq i,j \leq m-1)$ avec $I_{i,j}(\theta) = -\mathbb{E}_\theta\left[\dfrac{\partial^2 L_1(X_1;\theta)}{\partial\theta_i \partial\theta_j}\right]$:

$$I_{i,j}(\theta) = \frac{\mathbb{P}_\theta(X_1 = a_i)}{\theta_i^2}\mathbf{1}_{\{i=j\}} + \frac{\mathbb{P}_\theta(X_1 = a_m)}{(1-\sum_{l=1}^{m-1}\theta_l)^2} = \frac{1}{\theta_i}\mathbf{1}_{\{i=j\}} + \frac{1}{1-\sum_{l=1}^{m-1}\theta_l}\,.$$

L'information de Fisher $I(\theta)$ est la somme de la matrice diagonale $\Delta(\theta) = \text{Diag}\left(\frac{1}{\theta_1},\dots,\frac{1}{\theta_{m-1}}\right)$ et de la matrice $\frac{1}{1-\sum_{l=1}^{m-1}\theta_l}\mathbf{1}$, où la matrice $\mathbf{1}$ de taille $(m-1)\times(m-1)$ n'est composée que de 1. On calcule l'estimateur du maximum de vraisemblance $\hat\theta_n$ de θ. La log-vraisemblance de l'échantillon de taille n est :

$$L_n(X_1,\dots,X_n;\theta) = \sum_{i=1}^{m-1} N_i \log(\theta_i) + N_m \log\left(1-\sum_{l=1}^{m-1}\theta_l\right),$$

où $N_i = \sum_{k=1}^n \mathbf{1}_{\{X_k=a_i\}}$ représente les occurrences de i dans l'échantillon de taille n. On vérifie que si pour tout $i\in\{1,\dots,m\}$, $N_i\neq 0$, alors la limite de la log-vraisemblance sur la frontière de Θ est $-\infty$. Pour trouver le maximum de la log-vraisemblance, il suffit donc d'annuler les dérivées premières :

$$0 = \frac{\partial L_n(X_1,\dots,X_n;\hat\theta)}{\partial\theta_i} = \frac{N_i}{\hat\theta_i} - \frac{N_m}{1-\sum_{l=1}^{m-1}\hat\theta_l}, \quad \text{pour} \quad 1\le i\le m-1.$$

On en déduit que $\frac{N_i}{\hat\theta_i} = \frac{N_m}{1-\sum_{l=1}^{m-1}\hat\theta_l} = c$ qui est donc une constante indépendante de i. En utilisant la relation $\sum_{i=1}^m N_i = n$, il vient :

$$n = \sum_{l=1}^{m-1} N_l + N_m = \frac{N_m}{1-\sum_{l=1}^{m-1}\hat\theta_l}\sum_{l=1}^{m-1}\hat\theta_l + N_m = \frac{N_m}{1-\sum_{l=1}^{m-1}\hat\theta_l} = c.$$

On en déduit que $\hat\theta_i = \frac{N_i}{n}$. L'estimateur du maximum de vraisemblance est :

$$\hat\theta = (\hat\theta_1,\dots,\hat\theta_{m-1}) = \left(\frac{N_1}{n},\dots,\frac{N_{m-1}}{n}\right).$$

On a en particulier démontré le résultat suivant.

Corollaire IX.32. *L'estimateur du maximum de vraisemblance du vecteur des fréquences $p=(p_1,\dots,p_m)$ est le vecteur des fréquences empiriques $\left(\frac{N_1}{n},\dots,\frac{N_m}{n}\right)$.*

Soit $p(\hat\gamma_n) = (p_1(\hat\gamma_n),\dots,p_m(\hat\gamma_n))$ l'estimateur du maximum de vraisemblance de p sous la contrainte $p=p(\gamma)$. On considère l'estimateur du maximum de vraisemblance $\theta(\hat\gamma_n) = (p_1(\hat\gamma_n),\dots,p_{m-1}(\hat\gamma_n))$ de θ contraint.

On remarque que si $v = (v_1, \ldots, v_{m-1})^t \in \mathbb{R}^{m-1}$, on a :

$$v^t I(\theta) v = v^t \Delta(\theta) v + v^t \frac{1}{1 - \sum_{l=1}^{m-1} \theta_l} \mathbf{1} v = \sum_{i=1}^{m-1} \frac{v_i^2}{\theta_i} + \frac{\left(\sum_{l=1}^{m-1} v_l\right)^2}{1 - \sum_{l=1}^{m-1} \theta_l}.$$

On en déduit que :

$$\begin{aligned} \zeta_n^{(1)} &= n(\hat{\theta} - \theta(\hat{\gamma}_n))^t I(\hat{\theta})(\hat{\theta} - \theta(\hat{\gamma}_n)) \\ &= n \left[\sum_{i=1}^{m-1} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{\hat{p}_i} + \frac{\left(\sum_{l=1}^{m-1} (\hat{p}_l - p_l(\hat{\gamma}_n))\right)^2}{1 - \sum_{l=1}^{m-1} \hat{p}_l} \right]. \end{aligned}$$

Comme $\sum_{l=1}^{m} \hat{p}_l = 1$ et $\sum_{l=1}^{m} p_l(\hat{\gamma}_n) = 1$, on en déduit que $\sum_{l=1}^{m-1} (\hat{p}_l - p_l(\hat{\gamma}_n)) = p_m(\hat{\gamma} n) - \hat{p}_m$. Donc on a $\zeta_n^{(1)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{\hat{p}_i}$ et de manière similaire $\zeta_n^{(2)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{p_i(\hat{\gamma}_n)}$. Les hypothèses de la proposition IX.30 sont vérifiées. Ainsi le couple $(\zeta_n^{(1)}, \zeta_n^{(2)})$ converge sous H_0 vers (Z, Z) où Z est de loi χ^2 à $m-1-q$ degrés de liberté et les statistiques de test $\zeta_n^{(1)}$ et $\zeta_n^{(2)}$ convergent sous H_1 en probabilité vers $+\infty$.

IX.10 Autres tests asymptotiques

On présente les tests non-paramétriques de Kolmogorov-Smirnov sur le test d'adéquation à une loi, paragraphe IX.10.1, et le test de comparaison entre deux échantillons, paragraphe IX.10.2. Le paragraphe IX.10.3 détaille le test de comparaison entre deux échantillons dans un cadre paramétrique.

IX.10.1 Test de Kolmogorov-Smirnov pour un échantillon

Dans le cas discret, on a vu au paragraphe IX.9.2 un test d'adéquation de loi construit à partir du test du χ^2. Cette méthode est utilisable pour des lois discrètes à support fini. Le test asymptotique d'adéquation non-paramétrique de Kolmogorov-Smirnov permet de répondre à cette question pour des variables aléatoires réelles.

1. Le modèle.

On considère des variables aléatoires $(X_n, n \in \mathbb{N}^*)$ à valeurs réelles indépendantes, de même loi et dont la fonction de répartition F est continue.

2. Les hypothèses.

Soit F^0 une fonction de répartition continue connue. On désire établir un test pour savoir si X_1 a pour fonction de répartition F^0 : $H_0 = \{F = F^0\}$ contre l'hypothèse alternative $H_1 = \{F \neq F^0\}$. Il ne s'agit plus d'un test paramétrique.

Remarque IX.33. On peut discrétiser les variables aléatoires pour se ramener au test d'adéquation du paragraphe IX.9.2. On se donne une partition de $\mathbb{R}$ en m sous-ensembles (mesurables) $I_1, \ldots, I_m$. On définit $Y_n = i$ si $X_n \in I_i$. Les variables aléatoires $(Y_n, n \in \mathbb{N}^*)$ sont indépendantes, de même loi et à valeurs dans $\{1, \ldots, m\}$. On a $p_i = \mathbb{P}(Y_n = i) = \mathbb{P}(X_1 \in I_i)$. On peut alors tester si $p = (p_1, \ldots, p_m)$ est égal à la valeur théorique $p^0 = (p_1^0, \ldots, p_m^0)$ avec $p_i^0 = \mathbb{P}(X \in I_i)$, où la variable aléatoire X a pour fonction de répartition F^0. On teste ainsi $H_0' = \{p = p^0\}$ à l'aide du test d'adéquation du χ^2. Si on rejette l'hypothèse nulle H_0', alors on rejette H_0. En revanche si on accepte H_0', on ne peut pas en conclure que X_1 a pour fonction de répartition F^0. Ce test partiel est toutefois très facile à mettre en œuvre et peut être utilisé pour des variables aléatoires vectorielles quelconques. ◊

3. Les statistiques de test.

On considère la fonction de **répartition empirique** de l'échantillon de taille n :

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq y\}}.$$

On considère la statistique de test :

$$\zeta_n = \sqrt{n} \sup_{y \in \mathbb{R}} \left| F_n(y) - F^0(y) \right|.$$

4-5. Comportement sous H_0 et H_1.

Le théorème de Glivenko-Cantelli V.27 assure que $\sup_{y \in \mathbb{R}} |F_n(y) - F(y)|$ converge p.s. vers 0. On en déduit que sous H_1 la statistique de test converge p.s. vers $+\infty$.

On étudie maintenant le comportement de la statistique de test sous H_0. Le lemme suivant assure que la loi de ζ_n sous H_0 ne dépend pas de F^0. C'est en fait ce résultat qui motive le choix de cette statistique de test.

Lemme IX.34. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. réelles indépendantes et de même loi de fonction de répartition F. On suppose que la fonction F est continue. Soit F_n la fonction de répartition empirique de $X_1, \dots, X_n$. La loi de la variable aléatoire $\sup_{y\in\mathbb{R}} |F_n(y) - F(y)|$ ne dépend pas de F (et donc pas de la loi de X_1).*

Démonstration. On considère $F^{-1}(z) = \inf\{y; F(y) > z\}$, l'inverse généralisé de F. Cette fonction est bien définie car F est croissante. Si F est strictement croissante, alors F est une bijection, et F^{-1} correspond à l'inverse de F. Dans tous les cas, la continuité de F implique que $F(F^{-1}(z)) = z$ pour tout $z \in]0,1[$. On a donc :

$$\sup_{y\in\mathbb{R}} |F_n(y) - F(y)| = \sup_{z\in]0,1[} \left|F_n(F^{-1}(z)) - z\right|.$$

On remarque que $F_n(F^{-1}(z)) = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le F^{-1}(z)\}}$. Par définition, on a $\{y \le F^{-1}(z)\} = \{F(y) \le z\}$ pour tout $y \in \mathbb{R}$. Ainsi on a $F_n(F^{-1}(z)) = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\le z\}}$. C'est la fonction de répartition empirique de l'échantillon $(F(X_1), \dots, F(X_n))$. La fonction de répartition de $F(X_1)$ est, pour $z \in]0,1[$:

$$\mathbb{P}(F(X_1) \le z) = \mathbb{P}(X_1 \le F^{-1}(z)) = F(F^{-1}(z)) = z.$$

On reconnaît la fonction de répartition de la loi uniforme sur $[0,1]$. Les variables aléatoires $(F(X_n), n \in \mathbb{N}^*)$ sont indépendantes et de loi uniforme sur $[0,1]$. Ce dernier résultat est à rapprocher de la proposition III.22. On déduit donc de ce qui précède que $F_n \circ F^{-1}$ a même loi que la fonction de répartition empirique de n variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On en déduit ainsi que la loi de $\sup_{z\in]0,1[} \left|F_n(F^{-1}(z)) - z\right|$ est indépendante de la loi de X_1. □

On admet le théorème suivant.

Théorème IX.35 (Kolmogorov-Smirnov)**.** *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires réelles indépendantes de même loi et de fonction de répartition F continue. On a alors :*

$$\sqrt{n} \sup_{y\in\mathbb{R}} |F_n(y) - F(y)| \xrightarrow[n\to\infty]{\textbf{\textit{en loi}}} W,$$

où la fonction de répartition de la loi de W est définie, pour $w > 0$, par $\mathbb{P}(W \le w) = \sum_{k=-\infty}^{+\infty} (-1)^k \, \mathrm{e}^{-2k^2w^2}$.

On remarque que dans le théorème précédent la loi de W ne dépend pas de F, ce qui est cohérent avec le lemme IX.34.

On déduit de ce théorème que sous $H_0 = \{F = F^0\}$, la statistique de test ζ_n converge en loi vers W.

6. La région critique.

La statistique de test prend de très grandes valeurs sous H_1 et sous H_0 des valeurs typiques de la loi de W. On considère donc des régions critiques de la forme $W_n = \{\zeta_n \geq a\}$.

Un raisonnement similaire à celui du paragraphe IX.8.2 permet d'obtenir les résultats suivants.

7. Contrôle de l'erreur de 1ère espèce.

Le test de région critique $W_n = \{\zeta_n \geq w_{1-\alpha}\}$, où $w_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de W est de niveau asymptotique (non-uniforme) α.

8. Contrôle de l'erreur de 2ème espèce.

Le test est convergent.

9. La p-valeur.

Soit ζ_n^{obs} la statistique de test ζ_n calculée avec les n observations. La p-valeur asymptotique est :

$$p\text{-val} = \lim_{n\to\infty} \mathbb{P}_{F^0}(\zeta_n \geq \zeta_n^{\text{obs}}) = \mathbb{P}(W \geq \zeta_n^{\text{obs}}).$$

IX.10.2 Test de Kolmogorov-Smirnov pour deux échantillons

On présente le test de Kolmogorov-Smirnov pour deux échantillons qui permet de tester si deux échantillons sont issus de la même loi réelle.

1. Soit $(X_n, n \in \mathbb{N}^*)$ et $(Y_n, n \in \mathbb{N}^*)$ deux suites de variables aléatoires réelles, indépendantes et de même loi. On note F (resp. G) la fonction de répartition de X_1 (resp. Y_1). On suppose F et G continues.

2. On souhaite vérifier si les variables aléatoires X_1 et Y_1 ont même loi autrement dit si $F = G$. On considère l'hypothèse nulle $H_0 = \{F = G\}$ et l'hypothèse alternative $H_1 = \{F \neq G\}$.

3. On suppose que l'on dispose d'une réalisation de l'échantillon $(X_i, 1 \leq i \leq n_1)$ de taille n_1 et de l'échantillon $(Y_i, 1 \leq i \leq n_2)$ de taille n_2. On considère les fonctions de répartition empirique :

$$F_{n_1}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}_{\{X_i \leq x\}} \quad \text{et} \quad G_{n_2}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\{Y_i \leq x\}}$$

et la statistique de test :

$$\zeta_{n_1,n_2} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sup_{x \in \mathbb{R}} |F_{n_1}(x) - G_{n_2}(x)| .$$

4-8. On admet le résultat suivant :

Proposition IX.36. *Soit $(X_n, n \in \mathbb{N}^*)$ et $(Y_n, n \in \mathbb{N}^*)$ deux suites de variables aléatoires indépendantes et de même loi de fonctions de répartition respectives F et G. On suppose F et G continues. Si $F = G$, alors on a :*

$$\zeta_{n_1,n_2} \xrightarrow[\min(n_1,n_2)\to\infty]{\textbf{\textit{en loi}}} W,$$

où la fonction de répartition de la loi de W est définie, pour $w > 0$, par $\mathbb{P}(W \leq w) = \sum_{k=-\infty}^{+\infty} (-1)^k \, \mathrm{e}^{-2k^2 w^2}$.

En reprenant les résultats du paragraphe IX.10.1 on obtient que l'on rejette l'hypothèse nulle H_0 au niveau asymptotique α si ζ_{n_1,n_2} est supérieur au quantile d'ordre $1-\alpha$ de la loi de W et que ce test est convergent (quand $\min(n_1, n_2)$ tend vers $+\infty$).

9. D'après (IX.8), la p-valeur asymptotique non-uniforme (car seulement (IX.6) est vérifié) de ce test est donnée par $p\text{-val} = \mathbb{P}(W \geq \zeta^{\text{obs}}_{n_1,n_2})$, où $\zeta^{\text{obs}}_{n_1,n_2}$ est la statistique de test ζ_{n_1,n_2} calculée avec les observations.

IX.10.3 Test de comparaison pour deux échantillons

Dans un cadre paramétrique, on peut utiliser le test de Wald ou le test de Hausman pour vérifier si deux échantillons sont issus de la même loi. Ces tests sont

en général plus puissants que le test non-paramétrique de Kolmogorov-Smirnov à deux échantillons, mais l'hypothèse alternative est bien plus contraignante que dans le cadre non-paramétrique.

Ce test d'adéquation de loi est une application directe des paragraphes IX.8.2 et IX.8.3 si les deux échantillons ont même taille. On donne également un test asymptotique quand les échantillons sont de taille différente.

1. Soit $(X_n, n \in \mathbb{N}^*)$ et $(Y_n, n \in \mathbb{N}^*)$ deux suites de variables aléatoires indépendantes et de même loi. On suppose que la loi de X_1 et la loi de Y_1 appartiennent à une famille paramétrique $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$, où Θ est un ouvert de $\mathbb{R}^p$. On suppose de plus que le modèle est régulier et identifiable, que l'information de Fisher est une fonction continue du paramètre, inversible et d'inverse continue.

2. On souhaite vérifier si les variables aléatoires X_1 et Y_1 ont même loi autrement dit si $\theta^{(1)} = \theta^{(2)}$, où $\theta^{(1)}$ et $\theta^{(2)}$ sont respectivement les paramètres des lois de X_1 et Y_1. On considère l'hypothèse nulle $H_0 = \{\theta^{(1)} = \theta^{(2)}\}$ et l'hypothèse alternative $H_1 = \{\theta^{(1)} \neq \theta^{(2)}\}$.

3. On suppose que l'on dispose d'une réalisation de l'échantillon $(X_i, 1 \leq i \leq n_1)$ de taille n_1 et de l'échantillon $(Y_i, 1 \leq i \leq n_2)$ de taille n_2.

Le cas $n_1 = n_2$.

On pose $n = n_1 = n_2$. Les variables aléatoires $(Z_n = (X_n, Y_n), n \in \mathbb{N}^*)$ sont indépendantes de même loi. Ce modèle a pour paramètre $\tau = (\theta^{(1)}, \theta^{(2)})$, il est régulier et identifiable. On peut alors utiliser le test de Wald avec la fonction $g(\tau) = \theta^{(1)} - \theta^{(2)}$ ou le test de Hausman avec la fonction $h(\theta) = (\theta, \theta)$. La densité du modèle est :

$$p_n(x, y; \tau) = \prod_{i=1}^{n} p(x_i; \theta^{(1)}) \prod_{j=1}^{n} p(y_j; \theta^{(2)}).$$

On a :

$$\frac{\partial^2 \log p(x, y; \tau)}{\partial \tau^2} = \begin{pmatrix} \frac{\partial^2 \log p(x;\theta^{(1)})}{\partial \theta^2} & 0 \\ 0 & \frac{\partial^2 \log p(x;\theta^{(2)})}{\partial \theta^2} \end{pmatrix}.$$

L'information de Fisher est donc :

$$I(\theta^{(1)}, \theta^{(2)}) = \begin{pmatrix} J(\theta^{(1)}) & 0 \\ 0 & J(\theta^{(2)}) \end{pmatrix},$$

où $J(\theta) = -\mathbb{E}_\theta \left[\dfrac{\partial^2 \log p(X_1; \theta)}{\partial \theta \partial \theta'} \right]$ est l'information de Fisher du modèle initial. On remarque que l'estimateur du maximum de vraisemblance de τ, $\hat{\tau}_n$, est, à cause

de la forme produit de la densité de Z_1, $(\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$, où $\hat{\theta}_n^{(1)}$ est l'estimateur du maximum de vraisemblance de θ pour l'échantillon $X_1, \ldots, X_n$ et $\hat{\theta}_n^{(2)}$, l'estimateur du maximum de vraisemblance de θ pour l'échantillon $Y_1, \ldots, Y_n$.

La matrice du test de Wald est :

$$\Sigma(\tau) = \frac{\partial g}{\partial \tau}(\tau) I(\tau)^{-1} \left(\frac{\partial g}{\partial \tau}(\tau)\right)^t = J(\theta^{(1)})^{-1} + J(\theta^{(2)})^{-1}.$$

En particulier, sous H_0, on a $\theta^{(1)} = \theta^{(2)} = \theta$ et $\Sigma(\tau) = 2J(\theta)^{-1}$. On considère la statistique de test associée au test de Wald :

$$\zeta_n = \frac{1}{2}\, n(\hat{\theta}_n^{(1)} - \hat{\theta}_n^{(2)})^t J(\hat{\theta}_n^*)(\hat{\theta}_n^{(1)} - \hat{\theta}_n^{(2)}),$$

où $\hat{\theta}_n^*$ est un estimateur de θ sous H_0 (par exemple, on peut choisir $\hat{\theta}_n^* = (\hat{\theta}_n^{(1)} + \hat{\theta}_n^{(2)})/2$). Il s'agit d'une variante de la statistique de test (IX.9) dans la mesure où l'on choisit la forme particulière de Σ sous H_0. Néanmoins, la proposition IX.28 reste vraie pour cette statistique et sa démonstration est similaire. On considère également les statistiques de test (IX.10) et (IX.11) associées au test de Hausman :

$$\begin{aligned}
\zeta_n^{(1)} &= n(\hat{\tau}_n - h(\hat{\theta}_n))^t I(\hat{\tau}_n)(\hat{\tau}_n - h(\hat{\theta}_n)) \\
&= n(\hat{\theta}_n^{(1)} - \hat{\theta}_n)^t J(\hat{\theta}_n^{(1)})(\hat{\theta}_n^{(1)} - \hat{\theta}_n) + n(\hat{\theta}_n^{(2)} - \hat{\theta}_n)^t J(\hat{\theta}_n^{(2)})(\hat{\theta}_n^{(2)} - \hat{\theta}_n), \\
\zeta_n^{(2)} &= n(\hat{\tau}_n - h(\hat{\theta}_n))^t I((\hat{\theta}_n, \hat{\theta}_n))(\hat{\tau}_n - h(\hat{\theta}_n)) \\
&= n(\hat{\theta}_n^{(1)} - \hat{\theta}_n)^t J(\hat{\theta}_n)(\hat{\theta}_n^{(1)} - \hat{\theta}_n) + n(\hat{\theta}_n^{(2)} - \hat{\theta}_n)^t J(\hat{\theta}_n)(\hat{\theta}_n^{(2)} - \hat{\theta}_n),
\end{aligned}$$

où $\hat{\theta}_n$ maximise la vraisemblance sous H_0 définie par $p_n(x, y; \theta) = \prod_{i=1}^n p(x_i; \theta)$ $\prod_{j=1}^n p(y_j; \theta)$.

4-8. On déduit des paragraphes IX.8.2 et IX.8.3 que les trois statistiques de test convergent sous H_0 vers un χ^2 à p degrés de liberté et convergent sous H_1 en probabilité vers $+\infty$. Pour chaque statistique de test, on rejette l'hypothèse nulle d'égalité des lois H_0 au niveau asymptotique α si la statistique de test est supérieure au quantile d'ordre $1 - \alpha$ du χ^2 à p degrés de liberté. Ces tests sont convergents.

9. D'après (IX.8), la p-valeur asymptotique non-uniforme du test associé à la statistique de ζ_n (resp. $\zeta_n^{(i)}$ pour $i \in \{1, 2\}$) est p-val $= \mathbb{P}(Z \geq z)$, où $z = \zeta_n^{\text{obs}}$ (resp. $z = \zeta_n^{(i),\text{obs}}$) est la statistique de test calculée avec les n observations et Z est de loi $\chi^2(p)$.

Le cas $n_1 \neq n_2$.

On considère les statistiques de test suivantes :

$$\zeta_{n_1,n_2} = \frac{n_1 n_2}{n_1+n_2} n(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}^{(2)}_{n_2})^t J(\hat{\theta}^*_{n_1,n_2})(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}^{(2)}_{n_2}), \tag{IX.13}$$

$$\begin{aligned}\zeta^{(1)}_{n_1,n_2} &= n_1(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}_{n_1,n_2})^t J(\hat{\theta}^{(1)}_{n_1})(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}_{n_1,n_2}) \\ &\quad + n_2(\hat{\theta}^{(2)}_{n_2} - \hat{\theta}_{n_1,n_2})^t J(\hat{\theta}^{(2)}_{n_2})(\hat{\theta}^{(2)}_{n_2} - \hat{\theta}_{n_1,n_2}),\end{aligned} \tag{IX.14}$$

$$\begin{aligned}\zeta^{(2)}_{n_1,n_2} &= n_1(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}_{n_1,n_2})^t J(\hat{\theta}_{n_1,n_2})(\hat{\theta}^{(1)}_{n_1} - \hat{\theta}_{n_1,n_2}) \\ &\quad + n_2(\hat{\theta}^{(2)}_{n_2} - \hat{\theta}_{n_1,n_2})^t J(\hat{\theta}_{n_1,n_2})(\hat{\theta}^{(2)}_{n_2} - \hat{\theta}_{n_1,n_2}),\end{aligned} \tag{IX.15}$$

où $\hat{\theta}^{(1)}_{n_1}$ et $\hat{\theta}^{(2)}_{n_2}$ sont les estimateurs du maximum de vraisemblance de $\theta^{(1)}$ et $\theta^{(2)}$ à partir des échantillons de taille n_1 et n_2, $\hat{\theta}^*_{n_1,n_2}$ est un estimateur de $\theta = \theta^{(1)} = \theta^{(2)}$ sous H_0 (par exemple, on peut choisir $\hat{\theta}^*_{n_1,n_2} = (n_1\hat{\theta}^{(1)}_{n_1} + n_2\hat{\theta}^{(2)}_{n_2})/(n_1+n_2)$), et $\hat{\theta}_{n_1,n_2}$ est l'estimateur du maximum de vraisemblance de θ sous H_0.

4-8. On admet le résultat suivant.

Proposition IX.37. *Quand* $\min(n_1,n_2)$ *tend vers* $+\infty$*, les statistiques de test définies par (IX.13), (IX.14) et (IX.15) convergent en loi sous* H_0 *vers un* χ^2 *à* p *degrés de liberté et convergent en probabilité sous* H_1 *vers* $+\infty$*.*

On déduit de cette proposition que pour chaque statistique de test, on rejette l'hypothèse nulle d'égalité des lois H_0 au niveau asymptotique α si la statistique de test est supérieure au quantile d'ordre $1-\alpha$ du χ^2 à p degrés de liberté. La proposition assure également que ces tests sont convergents.

9. La p-valeur asymptotique non-uniforme du test associé à la statistique de ζ_{n_1,n_2} (resp. $\zeta^{(i)}_{n_1,n_2}$ pour $i \in \{1,2\}$) est p-val $= \mathbb{P}(Z \geq z)$, où $z = \zeta^{\text{obs}}_{n_1,n_2}$ (resp. $z = \zeta^{(i),\text{obs}}_{n_1,n_2}$) est la statistique de test calculée avec les observations et Z est de loi $\chi^2(p)$.

IX.11 Résumé

Soit $X = (X_1,\dots,X_n)$ un modèle d'échantillonnage de taille n. On note $p_n(x;\theta)$ sa densité, où $x = (x_1,\dots,x_n)$ représente les observations et $\theta \in \Theta$ est le paramètre du modèle. Soit H_0 l'hypothèse **nulle** et H_1 l'hypothèse **alternative** (H_0 et H_1 sont deux sous-ensembles non vides et disjoints de Θ) et φ un test.

- $\varphi(x)$ représente la probabilité de rejeter H_0 quand on observe x.
- Un test est **pur** si φ est à valeurs dans $\{0,1\}$. Dans ce cas si $\varphi(x)=0$, on accepte H_0 et si $\varphi(x)=1$, on refuse H_0 pour accepter H_1. $W_n=\{x;\varphi(x)=1\}$ est la **région critique** ou **zone de rejet** du test.
- L'**erreur de 1ère espèce** qui consiste à refuser H_0 alors qu'elle est vraie, est $\mathbb{P}_\theta(W_n)$ pour un test pur de région critique W_n avec $\theta\in H_0$.
- L'**erreur de 2ème espèce** qui consiste à accepter H_0 alors qu'elle est fausse, est $1-\mathbb{P}_\theta(W_n)$ pour un test pur de région critique W_n avec $\theta\in H_1$.
- La **puissance** est $\mathbb{P}_\theta(W_n)$ pour un test pur de région critique W_n avec $\theta\in H_1$.
- Le **niveau** d'un test pur de région critique W_n est $\alpha=\sup_{\theta\in H_0}\mathbb{P}_\theta(W_n)$.
- Le **principe de Neyman** consiste à minorer *a priori* l'erreur de 1ère espèce. On cherche alors parmi les tests de niveau inférieur à un seuil fixé, celui qui est le plus puissant.
- Si on peut écrire la région critique sous la forme $\{g(\zeta_n(x))\geq c\}$ où g est une fonction mesurable à valeurs réelles (souvent $g=Id$, $g=-Id$ ou $g=|\cdot|$, avec Id la fonction identité) et $\zeta_n=\zeta_n(X_1,\dots,X_n)$ est une statistique, alors ζ_n est appelée **statistique de test**.
- Dans le cas particulier où la région critique est de la forme $\{\zeta_n\geq c\}$ (resp. $\{\zeta_n\leq c\}$), la ***p*-valeur** du test est définie par $\sup_{\theta\in H_0}\mathbb{P}_\theta(\zeta_n\geq\zeta_n^{\text{obs}})$ (resp. $\sup_{\theta\in H_0}\mathbb{P}_\theta(\zeta_n\leq\zeta_n^{\text{obs}})$), où $\zeta_n^{\text{obs}}=\zeta_n(x_1^{\text{obs}},\dots,x_n^{\text{obs}})$ est la statistique de test évaluées en les observations $(x_1^{\text{obs}},\dots,x_n^{\text{obs}})$.
- Une famille de tests purs de régions critiques $(W_n,n\in\mathbb{N}^*)$, où W_n est une région critique pour un échantillon de taille n, est de **niveau asymptotique** (resp. **niveau asymptotique uniforme**) α si $\sup_{\theta\in H_0}\lim_{n\to\infty}\mathbb{P}_\theta(W_n)=\alpha$ (resp. $\lim_{n\to\infty}\sup_{\theta\in H_0}\mathbb{P}_\theta(W_n)=\alpha$). Le test est **convergent** si pour tout $\theta\in H_1$, on a $\lim_{n\to\infty}\mathbb{P}_\theta(W_n)=1$.
- Dans le cas particulier où les régions critiques sont de la forme $\{\zeta_n\geq c\}$ (resp. $\{\zeta_n\leq c\}$), la ***p*-valeur asymptotique uniforme** du test est, quand elle existe, $\lim_{n\to+\infty}\sup_{\theta\in H_0}\mathbb{P}_\theta(\zeta_n\geq z)$ (resp. $\lim_{n\to+\infty}\sup_{\theta\in H_0}\mathbb{P}_\theta(\zeta_n\leq z)$), où $z=\zeta_{n_0}^{\text{obs}}$ est la statistique de test évaluée en les n_0 observations dont on dispose. Souvent on ne dispose que de la p-valeur asymptotique non uniforme : $\sup_{\theta\in H_0}\lim_{n\to+\infty}\mathbb{P}_\theta(\zeta_n\geq z)$ (resp. $\sup_{\theta\in H_0}\lim_{n\to+\infty}\mathbb{P}_\theta(\zeta_n\leq z)$).

Un test est décrit par les étapes suivantes :

1. Modélisation : le choix du modèle dépend du problème concret considéré.
2. Choix des hypothèses : il est guidé par le fait que l'on ne souhaite pas rejeter H_0 à tort (majoration *a priori* de l'erreur de 1ère espèce).
3. Choix d'une statistique de test.
4. Comportement (asymptotique ou non) de la statistique de test sous H_0.
5. Comportement (asymptotique ou non) de la statistique de test sous H_1.

6. Région critique du test : elle correspond aux valeurs aberrantes de la statistique de test sous H_0 et aux valeurs raisonnables sous H_1.
7. Contrôle de l'erreur de 1ère espèce : détermination à partir du point 4 de la région critique en fonction du niveau du test (valeur exacte, majoration, valeur asymptotique ou valeur asymptotique uniforme).
8. Contrôle de l'erreur de 2ème espèce : analyse à partir du point 5 de la puissance ou convergence (approche asymptotique) du test.
9. Détermination de la p-valeur (valeur exacte, majoration, valeur asymptotique ou valeur asymptotique uniforme).

La p-valeur permet de conclure : **On rejette H_0 quand la p-valeur est "faible"**. La notion "faible" dépend du contexte ; en général "faible" signifie inférieur à quelques %.

Test de Neyman.

- 1-2. $H_0 = \{\theta_0\}$ et $H_1 = \{\theta_1\}$ sont des **hypothèses simples**.
- 3. La statistique de test est le **rapport de vraisemblance** $Z_n = \dfrac{p_n(X;\theta_1)}{p_n(X;\theta_0)}$.
- 4-8. Soit $\alpha \in]0,1[$. S'il existe κ tel que $\mathbb{P}_{\theta_0}(Z_n \geq \kappa) = \alpha$, alors le test de Neyman de région critique $W_n = \{Z_n \geq \kappa\}$ est un test uniformément plus puissant (UPP) de niveau α. (Si Z_n est sous $\mathbb{P}_{\theta_0}$ une v.a. continue, alors κ existe pour toute valeur de α.)
- 9. La p-valeur est $\mathbb{P}_{\theta_0}(Z_n \geq z_n^{\text{obs}})$, où z_n^{obs} est la statistique de test Z_n calculée en les observations.

- Sous certaines hypothèses, on peut établir pour les **modèles exponentiels** que le test précédent est un test UPP pour des **hypothèses unilatérales** : $H_0 = \{\theta < \theta_0\}$ ou $H_0 = \{\theta = \theta_0\}$ et $H_1 = \{\theta \geq \theta_1\}$ avec $\theta_0 \leq \theta_1$. Le rapport de vraisemblance s'exprime aisément à partir de la **statistique canonique** ζ_n du modèle exponentiel. Si la fonction de répartition de ζ_n sous $\mathbb{P}_{\theta_0}$ est continue, alors la région critique de niveau α est de la forme $W_n = \{\zeta_n \geq c\}$ ou $W_n = \{\zeta_n \leq c\}$, avec c tel que $\mathbb{P}_{\theta_0}(W_n) = \alpha$. La p-valeur est respectivement $\mathbb{P}_{\theta_0}(\zeta_n \geq \zeta_n^{\text{obs}})$ ou $\mathbb{P}_{\theta_0}(\zeta_n \leq \zeta_n^{\text{obs}})$ avec ζ_n^{obs} la statistique de test calculée en les observations.
- Sous certaines hypothèses, on peut établir pour les **modèles exponentiels** des tests uniformément plus puissants sans biais (UPPS) pour des **hypothèses bilatérales** : $H_0 = \{\theta \in [\theta_0, \theta_1]\}$ contre $H_1 = \{\theta \notin [\theta_0, \theta_1]\}$ avec $\theta_0 \leq \theta_1$. Si le test est pur, il a pour région critique $W_n = \{\zeta_n \notin]c_1, c_2[\}$ où c_1 et c_2 sont tels que $\mathbb{P}_{\theta_0}(W_n) = \mathbb{P}_{\theta_1}(W_n) = \alpha$.

Régression linéaire.

- 1. Le modèle est $X_i = \beta_0 + \sum_{k=1}^{p} \beta_k R_i^k + \varepsilon_i, \quad i \in \{1, \dots, n\}$, où R_i^k est le k-ième régresseur (non aléatoire) et $(\varepsilon_i, 1 \leq i \leq n)$ sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. Les paramètres $\beta_0, \dots, \beta_p$ et σ^2 sont inconnus.
- 2. On teste l'utilité de certains régresseurs (i.e. si les coefficients β_k sont nuls pour certaines valeurs de k).
- 3-9. Les résultats sont présentés dans la **table d'analyse de la variance** IX.1.

Test de Wald (hypothèses implicites).

- 1-2. On considère l'hypothèse nulle $H_0 = \{\theta \in \Theta; g(\theta) = 0\}$ et l'hypothèse alternative $H_1 = \{\theta \in \Theta; g(\theta) \neq 0\}$, où g est une fonction définie sur l'ouvert Θ et à valeurs dans $\mathbb{R}^r$ (et telle que $\partial g / \partial \theta$ soit de rang r).
- 3. La statistique de test est :

$$\zeta_n = n\, g(\hat{\theta}_n)^t \Sigma(\hat{\theta}_n)^{-1} g(\hat{\theta}_n) \quad \text{avec} \quad \Sigma(\theta) = \frac{\partial g}{\partial \theta}(\theta) I(\theta)^{-1} \left(\frac{\partial g}{\partial \theta}(\theta) \right)^t,$$

  où $I(\theta)$ est l'information de Fisher du modèle et $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ pour un échantillon de taille n.
- 4-8. Le test de région critique $W_n = \{\zeta_n \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(r)$ est de niveau asymptotique (non-uniforme) α. Ce test est convergent.
- 9. La p-valeur asymptotique est $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$, où Z est de loi $\chi^2(r)$ et ζ_n^{obs} est la statistique de test calculée en les observations.

Test de Hausman (hypothèses explicites).

- 1-2. On considère l'hypothèse nulle $H_0 = \{\theta = h(\gamma), \gamma \in \Gamma\}$ et l'hypothèse alternative $H_1 = \{\theta \notin h(\Gamma)\}$, où h est une fonction définie sur l'ouvert $\Gamma \subset \mathbb{R}^q$ à valeurs dans l'ouvert $\Theta \subset \mathbb{R}^p$ avec $q < p$ (et telle que $\partial h / \partial \gamma$ soit de rang q).
- 3. On considère deux statistiques de test :

$$\zeta_n^{(1)} = n(\hat{\theta}_n - h(\hat{\gamma}_n))^t I(\hat{\theta}_n)(\hat{\theta}_n - h(\hat{\gamma}_n)),$$
$$\zeta_n^{(2)} = n(\hat{\theta}_n - h(\hat{\gamma}_n))^t I(h(\hat{\gamma}_n))(\hat{\theta}_n - h(\hat{\gamma}_n)),$$

  où $I(\theta)$ est l'information de Fisher du modèle et $\hat{\theta}_n$ et $\hat{\gamma}_n$ sont les estimateurs du maximum de vraisemblance de θ et γ pour un échantillon de taille n.
- 4-8. Pour $i \in \{1, 2\}$, le test de région critique $W_n^{(i)} = \{\zeta_n^{(i)} \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(p - q)$ est de niveau asymptotique (non-uniforme) α. Ces tests sont convergents.

– 9. Les p-valeurs asymptotiques sont $\mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où Z est de loi $\chi^2(p-q)$ et $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée en les observations.

Test du χ^2 empirique.

– 1. On considère des variables aléatoires $(X_n, n \in \mathbb{N}^*)$ indépendantes de même loi et à valeurs dans $\{a_1, , \ldots, a_m\}$. On note $p = (p_1, \ldots, p_m)$ le vecteur des **fréquences** associé : $p_i = \mathbb{P}(X_1 = a_i)$.

– 2. On considère l'hypothèse nulle explicite $H_0 = \{p = h(\gamma); \gamma \in \Gamma\}$ et l'hypothèse alternative $H_1 = \{p \notin h(\Gamma)\}$ où h est définie sur un ouvert Γ de $\mathbb{R}^q$ avec $q < m-1$.

– 3. On considère les statistiques de test :

$$\zeta_n^{(1)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{\hat{p}_i} \quad \text{et} \quad \zeta_n^{(2)} = n \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i(\hat{\gamma}_n))^2}{p_i(\hat{\gamma}_n)},$$

où le vecteur des **fréquences empiriques** $\hat{p} = \left(\frac{N_1}{n}, \ldots, \frac{N_m}{n}\right)$ avec les occurrences $N_i = \sum_{k=1}^{n} \mathbf{1}_{\{X_k = i\}}$ est l'estimateur du maximum de vraisemblance de p et $\hat{\gamma}_n$ est l'estimateur du maximum de vraisemblance de γ.

– 4-8. Pour $i \in \{1, 2\}$, le test de région critique $W_n^{(i)} = \{\zeta_n^{(i)} \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(m-1-q)$ est de niveau asymptotique (non-uniforme) α. Ces tests sont convergents.

– 9. Les p-valeurs asymptotiques sont $\mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où Z est de loi $\chi^2(m-1-q)$ et $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée en les observations.

– Pour utiliser ce test en pratique, on recommande de s'assurer que soit $N_i \geq 5$ ou 10 pour tout i soit $np_i \geq 5$ ou 10 pour tout i.

Test d'adéquation à une loi (cas fini).

– 1. Même modèle que pour le test du χ^2 empirique.

– 2. On considère l'hypothèse nulle $H_0 = \left\{p = p^0\right\}$ et l'hypothèse alternative $H_1 = \{p \neq p^0\}$.

– 3. On considère les statistiques de test :

$$\zeta_n^{(1)} = n \sum_{i=1}^{m} \frac{\left(\hat{p}_i - p_i^0\right)^2}{\hat{p}_i} \quad \text{et} \quad \zeta_n^{(2)} = n \sum_{i=1}^{m} \frac{\left(\hat{p}_i - p_i^0\right)^2}{p_i^0}.$$

– 4-8. Pour $i \in \{1, 2\}$, le test de région critique $W_n^{(i)} = \{\zeta_n^{(i)} \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(m-1)$ est de niveau asymptotique (non-uniforme) α. Ces tests sont convergents.

– 9. Les p-valeurs asymptotiques sont $\mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où Z est de loi $\chi^2(m-1)$ et $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée en les observations.

Test d'indépendance.

- 1. Soit $(X_n = (V_n, W_n), n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes à valeurs dans $\{b_1, \dots, b_\ell\} \times \{c_1, \dots, c_r\}$.
- 2. On considère l'hypothèse nulle d'indépendance $H_0 = \{V_1 \text{ et } W_1 \text{ sont indépendants}\}$ et son alternative $H_1 = \{V_1 \text{ et } W_1 \text{ ne sont pas indépendants}\}$.
- 3. Les statistiques de test sont :

$$\zeta_n^{(1)} = \sum_{i=1}^{\ell} \sum_{j=1}^{r} \frac{\left(N_{i,j} - \frac{N_{i,\cdot} N_{\cdot,j}}{n}\right)^2}{N_{i,j}},$$

$$\zeta_n^{(2)} = n \sum_{i=1}^{\ell} \sum_{j=1}^{r} \frac{\left(N_{i,j} - \frac{N_{i,\cdot} N_{\cdot,j}}{n}\right)^2}{N_{i,\cdot} N_{\cdot,j}}.$$

- 4-8. Pour $i \in \{1, 2\}$, le test de région critique $W_n^{(i)} = \{\zeta_n^{(i)} \geq z_{1-\alpha}\}$, où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2((\ell-1)(r-1))$ est de niveau asymptotique (non-uniforme) α. Ces tests sont convergents.
- 9. Les p-valeurs asymptotiques sont $\mathbb{P}(Z \geq \zeta_n^{(i),\text{obs}})$, où Z est de loi $\chi^2((\ell-1)(r-1))$ et $\zeta_n^{(i),\text{obs}}$ est la statistique de test $\zeta_n^{(i)}$ calculée en les observations.

Test de Kolmogorov-Smirnov ou test non paramétrique d'adéquation à une loi (cas continu).

- 1. On considère des variables aléatoires $(X_n, n \in \mathbb{N}^*)$ à valeurs réelles indépendantes, de même loi et dont la fonction de répartition F est continue.
- 2. On considère l'hypothèse nulle d'adéquation à une loi particulière de fonction de répartition continue F^0 : $H_0 = \{F = F^0\}$ et l'hypothèse alternative $H_1 = \{F \neq F^0\}$.
- 3. La statistique de test est :

$$\zeta_n = \sqrt{n} \sup_{y \in \mathbb{R}} \left|F_n(y) - F^0(y)\right|,$$

où $F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq y\}}$ est la fonction de **répartition empirique**.

- 4-8. Le test de région critique $W_n = \{\zeta_n \geq w_{1-\alpha}\}$, où $w_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de W de fonction de répartition $\mathbb{P}(W \leq w) = \sum_{k=-\infty}^{+\infty} (-1)^k \, \mathrm{e}^{-2k^2 w^2}$ est de niveau asymptotique (non-uniforme) α. Ce test est convergent.
- 9. La p-valeur asymptotique est $\mathbb{P}(W \geq \zeta_n^{\text{obs}})$, où ζ_n^{obs} est la statistique de test calculée en les observations.

Test de Kolmogorov-Smirnov pour deux échantillons.

- 1. On considère $(X_n, n \in \mathbb{N}^*)$ et $(Y_n, n \in \mathbb{N}^*)$ deux suites indépendantes de variables aléatoires indépendantes et de même loi de fonctions de répartition respectives F et G continues.
- 2. On considère l'hypothèse nulle d'égalité des lois $H_0 = \{F = G\}$ et l'hypothèse alternative $H_1 = \{F \neq G\}$.
- 3. La statistique de test est :

$$\zeta_{n_1,n_2} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sup_{x \in \mathbb{R}} |F_{n_1}(x) - G_{n_2}(x)|,$$

où $F_{n_1}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}_{\{X_i \leq x\}}$ et $G_{n_2}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\{Y_i \leq x\}}$ sont les fonctions de répartition empiriques de l'échantillon $(X_i, 1 \leq i \leq n_1)$ de taille n_1 et de l'échantillon $(Y_i, 1 \leq i \leq n_2)$ de taille n_2.

- 4-8. Le test de région critique $W_{n_1,n_2} = \{\zeta_{n_1,n_2} \geq w_{1-\alpha}\}$, où $w_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de W définie dans le test de Kolmogorov-Smirnov, est de niveau asymptotique (non-uniforme) α. Ce test est convergent.
- 9. La p-valeur asymptotique est $\mathbb{P}(W \geq \zeta^{\text{obs}}_{n_1,n_2})$, où $\zeta^{\text{obs}}_{n_1,n_2}$ est la statistique de test calculée en les observations.

Test de comparaison pour deux échantillons.
Si on est dans un cadre paramétrique, on peut utiliser un test de Wald ou un test de Hausman plutôt que le test de Kolmogorov-Smirnov pour deux échantillons.

IX.12 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice IX.1 p. 250, *Exercice* IX.2 p. 264, *Exercice* IX.3 p. 264, *Exercice* IX.4 p. 267, *Exercice* IX.5 p. 269, *Exercice* IX.6 p. 272.

Exercice IX.7.
Le Docteur March a quatre filles. Est-il raisonnable de supposer que leur prochain enfant a une chance sur deux d'être un garçon ? Donner la p-valeur du test. △

Exercice IX.8.
Soit $X_1, \ldots, X_n$ un n-échantillon de loi exponentielle de paramètre $1/\theta$.

1. Construire le test de niveau α $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$.
2. Construire le test de niveau α $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta \neq \theta_0\}$.

△

Exercice IX.9.
Le tableau IX.6 fournit des données concernant l'âge (R^1), le kilométrage en milliers de kms (R^2), et le prix en milliers d'euros (X) pour un échantillon de voitures d'occasion d'un même type. On souhaite étudier la dépendance du prix en fonction des deux régresseurs de l'âge et du kilométrage. On considère le modèle de régression linéaire :

$$X_i = \beta_0 + \beta_1 R_i^1 + \beta_2 R_i^2 + \varepsilon_i,$$

où les variables aléatoires $(\varepsilon_i, i \in \mathbb{N}^*)$ sont indépendantes de même loi gaussienne $\mathcal{N}(0, \sigma^2)$.

R^1	5	4	6	5	5	5	6	6	2	7	7
R^2	92	64	124	97	79	76	93	63	13	111	143
X	7.8	9.5	6.4	7.5	8.1	9.0	6.1	8.7	15.4	6.4	4.4

Table IX.6. Âge (R^1), kilométrage en milliers de kms (R^2) et prix en millier d'euro (X) de plusieurs voitures d'occasion.

1. Visualiser les nuages de points (R^1, X) et (R^2, X). Pourquoi le modèle de régression linéaire semble-t-il pertinent ?
2. Calculer les estimateurs des paramètres $\beta_0, \beta_1, \beta_2$ et σ^2.
3. L'âge de la voiture a-t-il une influence significative (au niveau de 5%) sur le prix ?

4. Le kilométrage de la voiture a-t-il une influence significative (au niveau de 5%) sur le prix ?
5. En considérant les p-valeurs associées aux deux questions précédentes, quel modèle de régression linéaire simple (prix en fonction de l'âge ou prix en fonction du kilométrage) explique le mieux le prix d'une voiture ?

△

Exercice IX.10.
On compte en 1996 environ 3,9 millions de naissances aux USA qui se décomposent en 1 990 480 garçons et 1 901 014 filles. On désire savoir si l'on peut affirmer qu'il naît autant de filles que de garçons.

1. Construire un modèle probabiliste, à l'aide de variables aléatoires de Bernoulli, qui rend compte du problème considéré.
2. Vérifier qu'il s'agit d'un modèle exponentiel. En déduire un test pour tester l'hypothèse nulle $H_0 = \{$il naît autant de filles que de garçons$\}$ contre son alternative $H_1 = \{$il ne naît pas autant de filles que de garçons$\}$ ou bien l'alternative plus naturelle $H_1' = \{$il naît moins de filles que de garçons$\}$. Conclure en donnant la p-valeur.
3. Utiliser un test (asymptotique) d'adéquation de loi. Conclure en donnant la p-valeur.
4. Comparer ce test avec un test de Wald.
5. Donner, à l'aide d'un intervalle de confiance, une estimation de la probabilité qu'un nouveau-né soit un garçon. Comparer ce résultat avec les tests précédents.

△

Exercice IX.11.
Les dés de Weldon. Weldon a effectué $n = 26306$ lancers de douze dés à six faces[2]. On note X_i le nombre de faces indiquant cinq ou six lors du i-ième lancer. Les fréquences empiriques observées sont notées :

$$\hat{f}_j = \frac{N_j}{n} \, ,$$

où N_j est le nombre de fois où l'on a observé j faces indiquant cinq ou six, sur les douze lancers $N_j = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = j\}}$. Les observations sont données dans les tableaux IX.7 et IX.8.

2. W. Feller, *An introduction to probability theory and its applications*, volume 1, third ed., p. 148. Wiley, 1968.

$N_0 = 185$	$N_1 = 1149$	$N_2 = 3265$	$N_3 = 5475$
$N_4 = 6114$	$N_5 = 5194$	$N_6 = 3067$	$N_7 = 1331$
$N_8 = 403$	$N_9 = 105$	$N_{10} = 14$	$N_{11} = 4$
$N_{12} = 0$			

Table IX.7. Observations

$\hat{f}_0 = 0.007033$	$\hat{f}_1 = 0.043678$	$\hat{f}_2 = 0.124116$	$\hat{f}_3 = 0.208127$
$\hat{f}_4 = 0.232418$	$\hat{f}_5 = 0.197445$	$\hat{f}_6 = 0.116589$	$\hat{f}_7 = 0.050597$
$\hat{f}_8 = 0.015320$	$\hat{f}_9 = 0.003991$	$\hat{f}_{10} = 0.000532$	$\hat{f}_{11} = 0.000152$
$\hat{f}_{12} = 0.000000$			

Table IX.8. Fréquences empiriques observées

Si les dés sont non biaisés, la probabilité d'observer les faces cinq ou six dans un lancer de dés est de $1/3$. Les variables aléatoires $(X_i, 1 \leq i \leq n)$ suivent donc la loi binomiale de paramètres 12 et $1/3$. Les fréquences théoriques sont données dans le tableau IX.9.

$f_0 = 0.007707$	$f_1 = 0.046244$	$f_2 = 0.127171$	$f_3 = 0.211952$
$f_4 = 0.238446$	$f_5 = 0.190757$	$f_6 = 0.111275$	$f_7 = 0.047689$
$f_8 = 0.014903$	$f_9 = 0.003312$	$f_{10} = 0.000497$	$f_{11} = 0.000045$
$f_{12} = 0.000002$			

Table IX.9. Fréquences théoriques

1. Donner la statistique du test du χ^2 et la p-valeur. En déduire que l'on rejette l'hypothèse des dés non biaisés.
2. Rejette-t-on également l'hypothèse selon laquelle les variables sont distribuées suivant une loi binomiale de même paramètre $(12, r)$, r étant inconnu ?

△

Exercice IX.12.
Des sociologues s'intéressent à l'anxiété qui découle de la maladie. Des sociétés où il existe une explication orale de la maladie ("c'est parce qu'on a mangé du poison qu'on est malade", "c'est parce qu'on a reçu un sort", etc...) sont-elles différentes de celles où aucune explication orale n'existe ? Le tableau [3] IX.10 donne les notes d'anxiété qui ont été attribuées à chacune de ces sociétés ($n_1 = 16$ sociétés sans explication orale et $n_2 = 23$ sociétés avec explications orales).

3. S. Siegel, *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, 1956.

Sociétés sans expl. orale	Notes d'anxiété	Sociétés avec expl. orales	Notes d'anxiété
Lapp	13	Marquesans	17
Chamorro	12	Dobuans	16
Samoans	12	Baiga	15
Arapesh	10	Kwoma	15
Balinese	10	Thonga	15
Hopi	10	Alorese	14
Tanala	10	Chagga	14
Paiute	9	Navaho	14
Chenchu	8	Dahomeans	13
Teton	8	Lesu	13
Flathead	7	Masai	13
Papago	7	Lepeha	12
Wenda	7	Maori	12
Warrau	7	Pukapukans	12
Wogeo	7	Trobianders	12
Ontong	6	Kwakiull	11
		Manus	11
		Chiricahua	10
		Comanche	10
		Siriono	10
		Bena	8
		Slave	8
		Kurtatchi	6

Table IX.10. Angoisse dans les sociétés primitives.

Utiliser le test non-paramétrique de Kolmogorov-Smirnov pour tester l'hypothèse nulle $H_0 = \{$les deux types de sociétés ne sont pas différentes$\}$ contre l'hypothèse alternative contraire. Calculer la p-valeur asymptotique et conclure. (Comme les tailles des deux échantillons sont faibles, on peut également calculer la p-valeur par simulation.) △

X

Régions de confiance, Intervalles de confiance

On définit dans le paragraphe X.1 les régions de confiance qui généralisent la notion d'intervalle de confiance vue au paragraphe V.7 et dans l'exemple VI.14. Une région de confiance est donnée avec un niveau exact, voir le paragraphe X.1, ou approché (par excès ou asymptotique), voir le paragraphe X.2. Enfin, dans le paragraphe X.3 on montre les liens entre les régions de confiance et les tests.

Dans ce chapitre, on considère un modèle d'échantillonnage paramétrique $X_1, \ldots, X_n$ de taille n (cf. chapitre VIII.1). Les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et de même loi inconnue P, et elles sont à valeurs dans $\mathcal{X}$ ($\mathbb{R}$ ou $\mathbb{R}^d$). On suppose que la loi P appartient à une famille de probabilité $\mathcal{P} = \{\mathrm{P}_\theta; \theta \in \Theta\}$.

X.1 Régions et intervalles de confiance de niveau exact

Exemple. En reprenant l'exemple du fabricant de composants électroniques développé au chapitre VII, on remarque que la moyenne empirique $\hat{\theta}_n$ est l'estimateur efficace du paramètre inconnu θ. Il fournit une valeur approchée de θ, mais comme la loi de $\hat{\theta}_n$ est continue, on a $\mathbb{P}_\theta(\hat{\theta}_n = \theta) = 0$. Il est donc naturel de se donner un intervalle aléatoire Λ_n de $\Theta =]0, \infty[$ construit à partir de $\hat{\theta}_n$ tel que $\mathbb{P}_\theta(\theta \in \Lambda_n) = 1 - \alpha$, avec des valeurs typiques pour α : 5%, 1% ou 0,1%. Ce cas particulier est développé dans le chapitre VII.3. ◊

Définition X.1. *Soit g une application mesurable définie sur Θ réelle ou vectorielle et $x \mapsto \Lambda_n(x)$ une application de $\mathcal{X}^n$ à valeurs dans l'ensemble des boréliens de $g(\Theta)$. On dit que l'ensemble aléatoire $\Lambda_n = \Lambda_n(X_1, \ldots, X_n)$ est une* ***région de confiance*** *pour $g(\theta)$ de* ***niveau exact*** *$1 - \alpha$ si on a : pour tout $\theta \in \Theta$,*

$$\mathbb{P}_\theta\Big(g(\theta) \in \Lambda_n\Big) = 1 - \alpha.$$

Par simplicité, on omettra le mot exact. En général, on cherche à construire, pour un niveau donné, les régions de confiance les plus petites. Dans le cas où g est réelle ($g(\Theta) \subset \mathbb{R}$) et Λ_n est un intervalle, on parle d'**intervalle de confiance**.

Exemple X.2. On considère un échantillon de taille n d'un modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$. L'estimateur $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais efficace de μ. De plus la loi de $\bar{X}_n$ est la loi gaussienne $\mathcal{N}\left(\mu, \sigma_0^2/n\right)$ (cf. chapitre VI.2). Donc la variable aléatoire $\sqrt{n}(\bar{X}_n - \mu)/\sigma_0$ est de loi $\mathcal{N}(0,1)$. Soit des réels $z_- < z_+$ tels que $\displaystyle\int_{z_-}^{z_+} \mathrm{e}^{-y^2/2}\,\frac{dy}{\sqrt{2\pi}} = 1-\alpha$. On en déduit que $\mathbb{P}(\sqrt{n}(\bar{X}_n - \mu)/\sigma_0 \ \in [z_-, z_+]) = 1-\alpha$. On pose $I_n = \left[\bar{X}_n - \dfrac{\sigma_0 z_+}{\sqrt{n}}, \bar{X}_n - \dfrac{\sigma_0 z_-}{\sqrt{n}}\right]$. Comme :

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma_0 \ \in [z_-, z_+] \Longleftrightarrow \mu \in I_n,$$

il vient $\mathbb{P}_\mu(\mu \in I_n) = 1-\alpha$. Donc l'intervalle aléatoire I_n est un intervalle de confiance de μ de niveau $1-\alpha$. Il est facile de vérifier que minimiser la longueur de I_n, $\sigma_0(z_+ - z_-)/\sqrt{n}$, sous la contrainte $\displaystyle\int_{z_-}^{z_+} \mathrm{e}^{-y^2/2}\,\frac{dy}{\sqrt{2\pi}} = 1-\alpha$, revient à choisir $-z_- = z_+ = \phi_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. On choisit donc l'intervalle de confiance de μ de niveau $1-\alpha$:

$$\boxed{\left[\bar{X}_n - \frac{\sigma_0\phi_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{\sigma_0\phi_{1-\alpha/2}}{\sqrt{n}}\right].}$$

D'après les tables des quantiles de la loi gaussienne (cf. les paragraphes XI.1 et XI.2), les intervalles de confiance $\left[\bar{X}_n \pm \dfrac{1,96\ \sigma_0}{\sqrt{n}}\right]$ et $\left[\bar{X}_n \pm \dfrac{2,58\ \sigma_0}{\sqrt{n}}\right]$ sont respectivement de niveau 95% et 99%. ◊

Les fonctions pivotales définies ci-dessous permettent de construire facilement des régions de confiance.

Définition X.3. *Soit g une fonction mesurable définie sur Θ réelle ou vectorielle. Soit $\hat{g}_n = \hat{g}_n(X_1, \ldots, X_n)$ un estimateur de $g(\theta)$. Une fonction v réelle définie sur $g(\Theta)^2$ est **pivotale** pour $g(\theta)$ si la loi de $v(\hat{g}_n, g(\theta))$ est indépendante de θ.*

On peut alors construire une région de confiance pour $g(\theta)$ en choisissant un ensemble $B \subset \mathbb{R}$ tel que la quantité $\mathbb{P}_\theta\big(v(\hat{g}_n, g(\theta)) \in B\big)$, qui est indépendante de θ, soit égale à $1-\alpha$. On en déduit que $\Lambda_n(x) = \{x \in \mathcal{X}^n; v(\hat{g}_n(x), g(\theta)) \in B\}$ permet de définir une région de confiance de $g(\theta)$ de niveau $1-\alpha$.

Exemple. On reprend l'exemple développé au chapitre VII. On considère un échantillon de taille n dans un modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\lambda); \lambda > 0\}$. On désire estimer λ^{-1}. L'estimateur du maximum de vraisemblance $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais efficace de λ^{-1}. La loi de $n\bar{X}_n$ est une loi Gamma $\Gamma(\lambda, n)$. En particulier la loi de $\lambda n\bar{X}_n$ est $\Gamma(1, n)$. Elle est indépendante de λ. La fonction $v(\bar{X}_n, \lambda) = \lambda n\bar{X}_n$ est pivotale. On rappelle que la densité de la loi $\Gamma(1, n)$ est $f_n(y) = \dfrac{1}{\Gamma(n)}\, y^{n-1}\, \mathrm{e}^{-y}\, \mathbf{1}_{\{y>0\}}$. Soit $0 < a_n^- < a_n^+ < \infty$ tels que $\displaystyle\int_{a_n^-}^{a_n^+} f_n(y)\, dy = 1 - \alpha$. Il vient :

$$\mathbb{P}\left(\lambda^{-1} \in \left[\frac{n\bar{X}_n}{a_n^+}, \frac{n\bar{X}_n}{a_n^-}\right]\right) = \mathbb{P}(\lambda n\bar{X}_n \in [a_n^-, a_n^+]) = \int_{a_n^-}^{a_n^+} f_n(y)\, dy = 1 - \alpha.$$

On en déduit que $\left[\dfrac{n\bar{X}_n}{a_n^+}, \dfrac{n\bar{X}_n}{a_n^-}\right]$ est un intervalle de confiance de niveau $1 - \alpha$ pour λ^{-1}. Le choix de a_n^- et a_n^+ est arbitraire. Contrairement à l'exemple X.2, la longueur de l'intervalle de confiance est ici aléatoire. ♢

Exercice X.1.
On considère un échantillon de taille n dans un modèle gaussien $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\}$, où la moyenne et la variance sont inconnues. On pose $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ et $V_n = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2$.

1. Montrer que la loi de $\sqrt{n}(\bar{X}_n - \mu)/\sqrt{V_n}$ est indépendante de (μ, σ^2).
2. En déduire un intervalle de confiance symétrique de μ de niveau $1 - \alpha$.
3. Donner un intervalle approché pour n grand.

△

Correction. 1. On sait que le couple $(\bar{X}_n, V_n)$ est un estimateur sans biais optimal de (μ, σ^2) d'après les paragraphes VIII.4.4 et VIII.4.3, et le théorème VIII.33 du paragraphe VIII.5.3 sur les modèles exponentiels. De plus la proposition VI.12 assure que $\bar{X}_n$ et V_n sont indépendants. On déduit de la définition VI.13 que la loi de $\sqrt{n}\left(\bar{X}_n - \mu\right)/\sqrt{V_n}$ est une loi de Student de paramètre $n - 1$.

2. La densité de la loi de Student est symétrique par rapport à 0. Si $t_{(n-1),\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi de Student de paramètre $n - 1$, alors l'intervalle :

$$\boxed{\left[\bar{X}_n - \frac{t_{(n-1),\alpha/2}\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \frac{t_{(n-1),\alpha/2}\sqrt{V_n}}{\sqrt{n}}\right]}$$

est un intervalle de confiance de μ de niveau (exact) $1 - \alpha$, voir aussi l'exemple VI.14. Les valeurs des quantiles de la loi de Student de paramètre p sont tabulées.

3. Quand $p \to \infty$, d'après l'exercice VI.5, la loi de Student de paramètre p converge en loi vers la loi gaussienne $\mathcal{N}(0,1)$. Pour les grandes valeurs de p ($p \geq 20$) on peut utiliser la valeur des quantiles de la loi $\mathcal{N}(0,1)$: $t_{p,\alpha/2} \simeq \phi_{1-\alpha/2}$, où $\phi_{1-\alpha/2}$ est le quantile d'ordre $1-\dfrac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. L'intervalle $\left[\bar{X}_n - \dfrac{\phi_{1-\alpha/2}\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \dfrac{\phi_{1-\alpha/2}\sqrt{V_n}}{\sqrt{n}}\right]$ est un intervalle de confiance de μ de niveau asymptotique $1-\alpha$. ▲

Exercice X.2.
Déterminer dans l'exercice précédent un intervalle de confiance pour σ^2. △

Correction. La loi de $\dfrac{(n-1)V_n}{\sigma^2}$ est un χ^2 à $n-1$ degrés de liberté. Soit z_- et z_+ les quantiles d'ordre α_1 et $1-\alpha_2$ de la loi $\chi^2(n-1)$ avec $\alpha = \alpha_1 + \alpha_2$. On en déduit que l'intervalle $\left[\dfrac{(n-1)V_n}{z_+}, \dfrac{(n-1)V_n}{z_-}\right]$ est un intervalle de confiance pour σ^2 de niveau $1-\alpha$. ▲

Exercice X.3.
On considère un échantillon gaussien $\mathcal{P} = \left\{\mathcal{N}(\mu_0, \sigma^2); \sigma > 0\right\}$ où la moyenne μ_0 est connue. Rappeler l'estimateur du maximum de vraisemblance $\hat{\sigma}_n^2$ de σ^2. Vérifier que la loi de $\dfrac{\hat{\sigma}_n^2}{\sigma^2}$ est un χ^2 à n degrés de liberté (cf. l'exercice VI.3 du paragraphe VI.2). En déduire un intervalle de confiance pour σ^2 de niveau $1-\alpha$, et le comparer à celui obtenu dans l'exercice X.2. △

X.2 Régions et intervalles de confiance de niveau approché

X.2.1 Niveau par excès

On définit les régions de confiance de niveau par excès.

Définition X.4. *Soit g une application mesurable définie sur Θ réelle ou vectorielle et $x \mapsto \Lambda_n(x)$ une application de $\mathcal{X}^n$ à valeurs dans l'ensemble des boréliens de $g(\Theta)$. On dit que l'ensemble aléatoire $\Lambda_n = \Lambda_n(X_1, \ldots, X_n)$ est une région de confiance pour $g(\theta)$ de niveau par* ***excès*** $1-\alpha$ *si on a : pour tout $\theta \in \Theta$,*

$$\mathbb{P}_\theta\Big(g(\theta) \in \Lambda_n\Big) \geq 1-\alpha.$$

L'avantage des régions de confiance par excès réside dans la proposition suivante. On rappelle que $X = (X_1, \dots, X_n)$ représente le vecteur de l'échantillon de taille n.

Proposition X.5. *Soit $g(\theta) = (g_1(\theta), \dots, g_p(\theta))$. On suppose que pour tout $i \in \{1, \dots, p\}$, $\Lambda_n^{(i)}(X)$ est une région de confiance de niveau par excès $1 - \alpha_i$ pour $g_i(\theta)$. On suppose que $\sum_{i=1}^n \alpha_i < 1$. Alors $\Lambda_n(X) = \prod_{i=1}^n \Lambda_n^{(i)}(X)$ est une région de confiance pour $g(\theta)$ de niveau par excès $1 - \sum_{i=1}^n \alpha_i$.*

Démonstration. On a :

$$\begin{aligned}
\mathbb{P}_\theta\left(g(\theta) \in \Lambda_n(X)\right) &= \mathbb{P}_\theta\left(\forall i \in \{1, \dots, p\}; g_i(\theta) \in \Lambda_n^{(i)}(X)\right) \\
&= 1 - \mathbb{P}_\theta\left(\exists i \in \{1, \dots, p\} \quad \text{tel que} \quad g_i(\theta) \notin \Lambda_n^{(i)} i(X)\right) \\
&\geq 1 - \sum_{i=1}^n \mathbb{P}_\theta\left(g_i(\theta) \notin \Lambda_n^{(i)}(X)\right) \\
&\geq 1 - \sum_{i=1}^n \alpha_i.
\end{aligned}$$

□

Exemple. Suite des exercices X.1 et X.2. Soit $\alpha = \alpha_1 + \alpha_2 + \alpha_3$, $t_{\alpha_1/2}$ le quantile d'ordre $1 - \alpha_1/2$ de la loi de Student de paramètre $n - 1$, z_- le quantile d'ordre α_2 de la loi du χ^2 à $n - 1$ degrés de liberté et z_+ le quantile d'ordre $1 - \alpha_3$ de la loi du χ^2 à $n - 1$ degrés de liberté.

On déduit de la proposition X.5 et des exercices X.1 et X.2 que la région :

$$\left[\bar{X}_n - \frac{t_{\alpha_1/2}\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \frac{t_{\alpha_1/2}\sqrt{V_n}}{\sqrt{n}}\right] \times \left[\frac{(n-1)V_n}{z_+}, \frac{(n-1)V_n}{z_-}\right],$$

est une région de confiance de (μ, σ^2) de niveau par excès $1 - \alpha$. ◊

Exemple X.6. Soit $(X_1, \dots, X_n)$ un échantillon de loi de Bernoulli de paramètre $\theta \in]0, 1[$. On déduit de l'inégalité de Tchebychev que :

$$\mathbb{P}_\theta(\left|\bar{X}_n - \theta\right| \geq a) \leq \frac{\mathrm{Var}_\theta(\bar{X}_n)}{a^2} = \frac{\mathrm{Var}_\theta(X_1)}{na^2} = \frac{\theta(1-\theta)}{na^2} \leq \frac{1}{4na^2},$$

où $\bar{X}_n$ est la moyenne empirique. On en déduit que $\left[\bar{X}_n - \frac{a}{2\sqrt{n}}, \bar{X}_n + \frac{a}{2\sqrt{n}}\right]$ est un intervalle de confiance pour θ de niveau par excès $1 - 1/a^2$. ◊

X.2.2 Niveau asymptotique

Pour les échantillons de grande taille, on peut utiliser une approche asymptotique.

Définition X.7. *Soit g une application mesurable définie sur Θ réelle ou vectorielle et, pour $n \in \mathbb{N}^*$, $x \mapsto \Lambda_n(x)$ une application de $\mathcal{X}^n$ à valeurs dans l'ensemble des boréliens de $g(\Theta)$. On dit que la suite d'ensembles aléatoires $\Lambda_n = \Lambda_n(X_1, \ldots, X_n)$ est une région de confiance pour $g(\theta)$ de niveau* ***asymptotique*** *$1-\alpha$ si on a : pour tout $\theta \in \Theta$,*

$$\lim_{n\to\infty} \mathbb{P}_\theta\Big(g(\theta) \in \Lambda_n\Big) = 1-\alpha.$$

Exemple. Suite de l'exemple X.6. On rappelle que $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \big(\frac{1}{n}\sum_{i=1}^n X_i\big)^2$ est un estimateur (biaisé) convergent de $\sigma^2 = \mathrm{Var}(X_1)$. Dans cet exemple, on a $\hat{\sigma}_n^2 = \bar{X}_n(1-\bar{X}_n)$. On considère l'intervalle aléatoire :

$$I_n = \left[\bar{X}_n - \frac{\phi_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}, \bar{X}_n + \frac{\phi_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right],$$

où $\phi_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. On déduit du paragraphe V.7 que :

$$\mathbb{P}_\theta\left(\theta \in I_n\right) \underset{n\to\infty}{\longrightarrow} \int_{-\phi_{1-\alpha/2}}^{\phi_{1-\alpha/2}} \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}} = 1-\alpha.$$

Donc l'intervalle aléatoire I_n est un intervalle de confiance pour θ de niveau asymptotique $1-\alpha$. Cet intervalle de confiance asymptotique est de mauvaise qualité si np ou $n(1-p)$ sont faibles, comme le suggère la majoration du théorème de Berry-Esséen V.34. ◊

Soit $\hat{\theta}_n$ un estimateur convergent asymptotiquement normal de $\theta \in \Theta$, où Θ est un **ouvert** de $\mathbb{R}^p$. On suppose que la matrice de covariance asymptotique $\Sigma(\theta)$ est inversible pour tout $\theta \in \Theta$ et que l'application $\theta \mapsto \Sigma(\theta)^{-1}$ est continue. Le théorème de Slutsky implique que le vecteur aléatoire :

$$\Sigma(\hat{\theta}_n)^{-1/2}\sqrt{n}\ (\hat{\theta}_n - \theta)$$

converge en loi sous $\mathbb{P}_\theta$ vers la loi gaussienne $\mathcal{N}(0, I_p)$, où I_p est la matrice identité de taille $p \times p$. On en déduit que la variable aléatoire réelle :

$$n(\hat{\theta}_n - \theta)^t \Sigma(\hat{\theta}_n)^{-1}(\hat{\theta}_n - \theta)$$

converge en loi vers Z de loi χ^2 à p degrés de liberté. Soit l'ellipsoïde aléatoire de $\mathbb{R}^p$ défini par :

$$\Lambda_n = \left\{ q \in \mathbb{R}^p ; n(\hat{\theta}_n - q)^t \Sigma(\hat{\theta}_n)^{-1} (\hat{\theta}_n - q) \leq z_\alpha \right\}, \tag{X.1}$$

où z_α est le quantile d'ordre $1-\alpha$ du χ^2 à p degrés de liberté. On a, grâce à la proposition V.19, que pour tout $\theta \in \Theta$:

$$\mathbb{P}_\theta \left(\theta \in \Lambda_n \right) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(Z \leq z_\alpha) = 1 - \alpha.$$

Donc l'ellipsoïde Λ_n est une région de confiance pour θ de niveau asymptotique $1-\alpha$.

Remarque. Soit $\hat{\theta}_n^{(2)}$ un estimateur convergent de θ asymptotiquement normal de matrice de covariance $\Sigma_{(2)}(\theta)$ telle que pour tout $\theta \in \Theta$, la matrice $\Sigma(\theta) - \Sigma_{(2)}(\theta)$ est définie positive. Alors pour un niveau asymptotique donné $1-\alpha$, le volume de l'ellipsoïde de confiance défini par (X.1) est plus grand que le volume de l'ellipsoïde de confiance $\Lambda_n^{(2)}$ défini par (X.1) avec Σ remplacé par $\Sigma_{(2)}$ et $\hat{\theta}_n$ par $\hat{\theta}_n^{(2)}$. Il est donc naturel de choisir $\Lambda_n^{(2)}$, l'ellipsoïde de confiance de volume le plus petit. Ceci est en accord avec le fait que $\hat{\theta}_n^{(2)}$ est asymptotiquement préférable à $\hat{\theta}_n$. ◊

Afin de minimiser la taille des ellipsoïdes de confiance de niveau asymptotique donné, il est naturel de considérer des estimateurs asymptotiquement efficaces, voir la définition VIII.45.

Proposition X.8. *Soit $\hat{\theta}_n$ un estimateur asymptotiquement efficace de θ. On suppose que l'information de Fisher $\theta \mapsto I(\theta)$ définie sur $\Theta \subset \mathbb{R}^p$ est continue et d'inverse continu. Soit $r \leq p$ et $g : \Theta \to \mathbb{R}^r$ de classe C^1 telle que la matrice $\dfrac{\partial g}{\partial \theta}$ est de rang r. Alors l'ellipsoïde aléatoire :*

$$\Lambda_n = \Big\{ g \in \mathbb{R}^r ; \\ n \left(g(\hat{\theta}_n) - g \right)^t \left[\frac{\partial g}{\partial \theta}(\hat{\theta}_n) I(\hat{\theta}_n)^{-1} \left(\frac{\partial g}{\partial \theta}(\hat{\theta}_n) \right)^t \right]^{-1} \left(g(\hat{\theta}_n) - g \right) \leq z_\alpha \Big\}, \tag{X.2}$$

où z_α est le quantile d'ordre $1-\alpha$ du χ^2 à r degrés de liberté, est une région de confiance de niveau asymptotique $1-\alpha$ pour $g(\theta)$.

Démonstration. Le corollaire VIII.46 assure que l'estimateur de $g(\hat{\theta}_n)$ de $g(\theta)$ est asymptotiquement normal de variance asymptotique $\Sigma(\theta) = \frac{\partial g}{\partial \theta}(\hat{\theta}) I(\hat{\theta})^{-1} \left(\frac{\partial g}{\partial \theta}(\hat{\theta}) \right)^t$.

Donc les variables aléatoires $\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right)$ convergent en loi vers un vecteur gaussien G de loi $\mathcal{N}(0, \Sigma(\theta))$. On remarque que la matrice $\Sigma(\theta)$ est régulière car la matrice $\dfrac{\partial g}{\partial \theta}$ est de rang r. De plus l'application $\theta \mapsto \Sigma(\theta)$ est continue. On en déduit que l'application $\theta \mapsto \Sigma(\theta)^{-1/2}$ est continue. Le théorème de Slutsky implique que $\Sigma(\hat{\theta}_n)^{-1/2}\sqrt{n}\,(g(\hat{\theta}_n) - g(\theta))$ converge en loi vers $Y = \Sigma(\theta)^{-1/2}G$ de loi $\mathcal{N}(0, I_r)$ où I_r est la matrice identité de taille $r \times r$. Par conséquent, la variable aléatoire

$$Z_n = n(g(\hat{\theta}_n) - g(\theta))^t \Sigma(\hat{\theta}_n)^{-1}(g(\hat{\theta}_n) - g(\theta)) = \left|\Sigma(\hat{\theta}_n)^{-1/2}\sqrt{n}\,(g(\hat{\theta}_n) - g(\theta))\right|^2$$

converge en loi vers $|Y|^2$ c'est-à-dire un χ^2 à r degrés de liberté. En particulier, on a $\lim_{n\to+\infty} \mathbb{P}_\theta(Z_n \leq z_\alpha) = 1 - \alpha$, où z_α est le quantile d'ordre $1 - \alpha$ du χ^2 à r degrés de liberté.

Pour conclure, il suffit de remarquer que $Z_n \leq z_\alpha \iff g(\theta) \in \Lambda_n$, où Λ_n est défini par (X.2). □

Pour g égal à l'identité, on obtient le corollaire suivant.

Corollaire X.9. *Si $\hat{\theta}_n$ est un estimateur asymptotiquement efficace du paramètre θ de dimension p et si l'information de Fisher est une fonction continue du paramètre, alors l'ellipsoïde aléatoire :*

$$\Lambda_n = \left\{ q \in \mathbb{R}^p ; n\left(\hat{\theta}_n - q\right)^t I(\hat{\theta}_n)\left(\hat{\theta}_n - q\right) \leq z_\alpha \right\},$$

où z_α est le quantile d'ordre $1 - \alpha$ du χ^2 à p degrés de liberté, est une région de confiance pour θ de niveau asymptotique $1 - \alpha$.

Remarque. Il reste une question importante qui concerne la validité réelle de ces régions de confiance. En effet le niveau $1 - \alpha$ est le niveau asymptotique quand la taille de l'échantillon n tend vers $+\infty$. Pour une taille fixée, le niveau exact est parfois très différent de $1-\alpha$. Voir le paragraphe VII.3 pour un exemple numérique. On peut recourir à des simulations pour avoir une estimation du niveau exact. ◊

La construction d'intervalle de confiance asymptotique dépasse le cadre paramétrique. En effet soit $(X_i ; i \in \mathbb{N}^*)$ une suite de variables aléatoires réelles indépendantes et de même loi inconnue. On suppose que $\mathbb{E}[X_i^2] < \infty$. On désire obtenir une estimation de $\mu = \mathbb{E}[X_i]$, sans supposer que la loi des X_i appartient à une famille paramétrique de lois. On sait par la loi forte des grands nombres que

$\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ converge presque sûrement vers μ. De plus $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mathrm{Var}(X_1)}}$ converge en loi vers la loi gaussienne $\mathcal{N}(0,1)$ par le théorème central limite. On déduit du théorème de Slutsky que $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{V_n}}$, où $V_n = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, converge en loi vers la loi gaussienne $\mathcal{N}(0,1)$. On en déduit que l'intervalle :

$$\boxed{\left[\bar{X}_n - \frac{\phi_{1-\alpha/2}\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \frac{\phi_{1-\alpha/2}\sqrt{V_n}}{\sqrt{n}}\right],}$$

où $\phi_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi gaussienne $\mathcal{N}(0,1)$, est un intervalle de confiance de μ de niveau asymptotique $1-\alpha$. Dans le cas où X_1 est de loi gaussienne, on peut calculer le niveau exact de cet intervalle de confiance, voir l'exercice X.1.

X.3 Régions de confiance et tests

Dans ce qui suit on montre comment on peut construire un test à partir de régions de confiance et réciproquement.

Soit g une fonction mesurable définie sur Θ réelle ou vectorielle. Soit $x \mapsto \Lambda_n(x)$, où $x \in \mathcal{X}^n$, une région de confiance pour $g(\theta)$ de niveau $1-\alpha$. On définit pour $g \in g(\Theta)$ l'ensemble :

$$W_n^g = \{x \in \mathcal{X}^n; g \notin \Lambda_n(x)\}.$$

On a l'équivalence $x \notin W_n^g \iff g \in \Lambda_n(x)$. On rappelle que $X = (X_1, \ldots, X_n)$ représente le vecteur de l'échantillon de taille n. Pour tout $\theta \in \Theta$, on a :

$$1 - \alpha = \mathbb{P}_\theta(g(\theta) \in \Lambda_n(X)) = 1 - \mathbb{P}_\theta(X \in W_n^{g(\theta)}).$$

On considère l'hypothèse nulle $H_0 = \{\theta \in \Theta; g(\theta) = g\}$ et l'hypothèse alternative $H_1 = \Theta \backslash H_0$. Pour tout $\theta \in H_0$, l'erreur de 1ère espèce du test pur de région critique W_n^g est $\mathbb{P}_\theta(W_n^g) = \mathbb{P}_\theta(X \in W_n^{g(\theta)}) = \alpha$. L'ensemble W_n^g est donc une région critique de niveau α pour le test de l'hypothèse nulle H_0 contre son alternative H_1.

Réciproquement si pour tout g, W_n^g est une région critique de niveau α pour le test de l'hypothèse nulle $H_0 = \{\theta \in \Theta; g(\theta) = g\}$ contre son alternative, alors $\Lambda_n(x) = \{g; x \notin W_n^g\}$ définit une région de confiance de niveau $1-\alpha$ pour $g(\theta)$.

Cette équivalence entre régions de confiance et régions critiques permet de définir des régions de confiance associées aux tests UPP et UPPS. On définit ainsi des régions de confiance optimales sur la base de critères intrinsèques.

X.4 Résumé

On retiendra la construction
- des régions de confiance de **niveau exact** dans le modèle gaussien ;
- des régions de confiance de **niveau par excès** pour des paramètres vectoriels comme produit des intervalles de confiance de chacune des composantes du paramètre ;
- des régions de confiance de **niveau asymptotique** à l'aide du théorème central limite ou à partir d'estimateurs asymptotiquement normaux.

Enfin on peut mettre en bijection les tests et les régions de confiance.

X.5 Exercices

Les exercices dans la partie du cours sont aux pages suivantes :

Exercice X.1 p. 295, *Exercice* X.2 p. 296, *Exercice* X.3 p. 296.

Exercice X.4.
On considère un échantillon gaussien $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\right\}$ de taille n, où la moyenne et la variance sont inconnues.

1. Donner des intervalles de confiance de niveau exact donné pour μ et pour σ^2.
2. En déduire une région de confiance de niveau par excès $1-\alpha$ pour (μ, σ^2).
3. Par simulation évaluer le niveau exact de la région de confiance de la question précédente.
4. Donner une région de confiance de niveau exact $1-\alpha$ pour (μ, σ^2). Comparer avec la région de confiance de niveau par excès $1-\alpha$.

△

Exercice X.5.
On considère un échantillon $X = (X_1, \ldots, X_n)$ de variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in]0,1[$ inconnu. On considère la statistique de la moyenne empirique $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$. La moyenne empirique est un estimateur convergent de p. On désire donner plusieurs intervalles de confiance pour p de niveau exact ou approché $1-\alpha$.

1. Vérifier que $\mathrm{Var}_p(\bar{X}_n) \leq 1/4$. Déduire de l'inégalité de Tchebychev, un intervalle de confiance (symétrique) de niveau par excès $1-\alpha$ pour p.
2. Montrer que $\bar{X}_n(1-\bar{X}_n)$ est un estimateur convergent de $\sigma^2 = \mathrm{Var}_p(X_1)$. Déduire du théorème de la limite centrale un intervalle de confiance de niveau asymptotique $1-\alpha$ pour p.
3. On considère la fonction $g_n(x,p) = \dfrac{\sqrt{n}(x-p)}{\sqrt{p(1-p)}}$. Montrer que la suite $(g_n(\bar{X}_n, p), n \in \mathbb{N}^*)$ converge en loi vers une loi indépendante du paramètre p (la suite de fonction $(g_n, n \in \mathbb{N}^*)$ est asymptotiquement pivotale). En considérant l'inverse de $p \mapsto g_n(x,p)$, donner un intervalle de confiance de niveau asymptotique $1-\alpha$ pour p.
4. En utilisant la majoration $\mathrm{Var}_p(X_1) \leq 1/4$, déduire du théorème central limite un intervalle de confiance de niveau asymptotique par excès $1-\alpha$ pour p.

5. Trouver une fonction g telle que $g'(\mathbb{E}_p[X_1])^2 \operatorname{Var}_p(X_1)$ soit constant pour tout $p \in]0,1[$. Montrer que la suite $(\sqrt{n}(g(\bar{X}_n) - g(p)), n \in \mathbb{N}^*)$ converge en loi et déterminer la limite. En déduire un intervalle de confiance de niveau asymptotique $1-\alpha$ pour p (méthode de stabilisation de la variance).
6. Soit $1 > a > p$. Montrer, en utilisant l'inégalité de Markov, que $\mathbb{P}_p(\bar{X}_n > a) \leq \mathbb{E}_p\left[\mathrm{e}^{\lambda \bar{X}_n - \lambda a}\right]$ pour tout $\lambda > 0$. Calculer puis minimiser le second terme en λ. En déduire que :

$$\mathbb{P}_p(\bar{X}_n > a) \leq \mathrm{e}^{-n[a\log(a/p)+(1-a)\log((1-a)/(1-p))]}.$$

   Déduire de cette majoration une majoration de $\mathbb{P}_p(\bar{X}_n < a)$ pour $0 < a < p$. Construire un intervalle de confiance de niveau par excès $1-\alpha$ pour p.
7. Comparer, en calculant numériquement le niveau exact, les 6 intervalles de confiance des questions précédentes.
8. On suppose $\alpha = 5\%$.
   a) Déduire de la question 2 que l'intervalle $\left[\bar{X}_n \pm \dfrac{1}{\sqrt{n}}\right]$ est un intervalle de confiance de niveau asymptotique par excès de 95% pour p.
   b) Vérifier numériquement que l'intervalle $\left[\bar{X}_n \pm \left(\dfrac{1}{\sqrt{n}} + \dfrac{1}{2n}\right)\right]$ est un intervalle de confiance de niveau par excès de 95% pour p.
9. On suppose que l'on observe la réalisation $x_i = 0$ pour $1 \leq i \leq n$ c'est-à-dire $\bar{x}_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i = 0$.
   a) Donner les intervalles de confiance correspondants aux questions 1-6, en faisant un développement limité en n. Vérifier que les intervalles pertinents sont de la forme $[0, c/n]$, où la constante c dépend de α.
   b) Soit c une constante. Après avoir vérifié que le modèle est exponentiel, donner un test UPP de seuil α pour l'hypothèse nulle $H_0 = \{p \leq c/n\}$ contre son alternative $H_1 = \{p > c/n\}$.
   c) Comparer l'intervalle de confiance associé au test UPP et les intervalles de confiance précédents.

△

XI

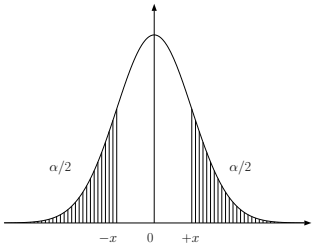
Tables statistiques

Les tables qui suivent ont été générées à l'aide du logiciel Scilab.

XI.1 Quantiles de la loi $\mathcal{N}(0,1)$

Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$. On a $\mathbb{P}(|X| \geq x) = 2\displaystyle\int_x^{+\infty} \mathrm{e}^{-y^2/2}\,\frac{dy}{\sqrt{2\pi}}$. On pose :

$$\mathbb{P}(|X| \geq x) = \alpha.$$



La table donne les valeurs de x en fonction de α, par exemple :

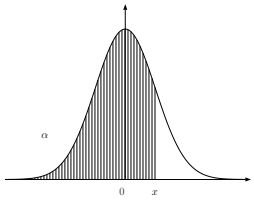
$\mathbb{P}(|X| \geq 0.6280) \simeq 0.53$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

XI.2 Fonction de répartition de la loi $\mathcal{N}(0,1)$

Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$. On a $\mathbb{P}(X \leq x) = \displaystyle\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\,\frac{dy}{\sqrt{2\pi}}$. On pose :

$$\mathbb{P}(X \leq x) = \alpha.$$



La table donne les valeurs de α en fonction de x, par exemple : $\mathbb{P}(X \leq 1.96) \simeq 0.97500$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

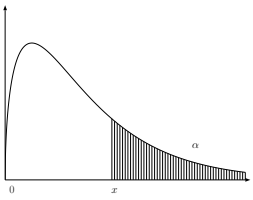
La table suivante donne $1-\alpha$ pour les grandes valeurs de x.

x	2	3	4	5	6	7	8	9	10
$1-\alpha$	2.28e-02	1.35e-03	3.17e-05	2.87e-07	9.87e-10	1.28e-12	6.22e-16	1.13e-19	7.62e-24

XI.3 Quantiles de la loi du χ^2

Soit X_n une variable aléatoire de loi $\chi^2(n)$. On a $\mathbb{P}(X_n \geq x) = \int_x^{+\infty} \frac{1}{2^{n/2}\Gamma(n/2)}\, y^{\frac{n}{2}-1}\, \mathrm{e}^{-y/2}\ dy$. On pose :

$$\mathbb{P}(X_n \geq x) = \alpha.$$



La table donne x en fonction de n et α, par exemple : $\mathbb{P}(X_8 \geq 20.09) \simeq 0.01$.

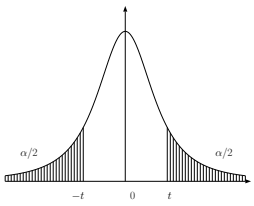
$n\backslash\alpha$	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.001
1	0.0002	0.0010	0.0039	0.0158	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.12
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	39.25
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.31
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	54.05
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

Lorsque $n > 30$, on peut utiliser l'approximation $\sqrt{2\chi^2(n)} - \sqrt{2n-1} \overset{\text{Loi}}{\simeq} \mathcal{N}(0,1)$ (voir l'exercice VI.13).

XI.4 Quantiles de la loi de Student

Soit X_n une variable aléatoire de loi de Student de paramètre n. On a :

$$\mathbb{P}(|X_n| \geq t) = 2\int_t^{+\infty} \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)}(1+\frac{y^2}{n})^{-(n+1)/2}\, dy.$$



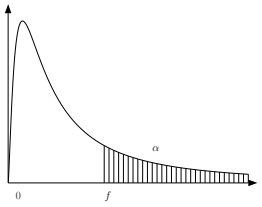
On pose $\mathbb{P}(|X_n| \geq t) = \alpha$. La table donne t en fonction de n et α, par exemple :
$\mathbb{P}(|X_{20}| \geq 2.086) \simeq 0.05$.

$n\backslash\alpha$	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100	0.050	0.020	0.010	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
80	0.126	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

XI.5 Quantiles de la loi de Fisher-Snedecor

Soit X de loi de Fisher-Snedecor de paramètre (n, m). On a :

$$\mathbb{P}(X \geq f) = \int_f^{+\infty} \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} \frac{y^{\frac{n}{2}-1}\, dy}{\left(1+\frac{yn}{m}\right)^{\frac{n+m}{2}}}.$$



On pose $\mathbb{P}(X_{n,m} \geq f) = \alpha$. La table donne f en fonction de n, m et α, par exemple : $\mathbb{P}(X_{4,20} \geq 4.43) \simeq 0.01$.

	$n=1$		$n=2$		$n=3$		$n=4$		$n=5$	
m	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01
1	161.45	4052.18	199.50	4999.50	215.71	5403.35	224.58	5624.58	230.16	5763.65
2	18.51	98.50	19.00	99.00	19.16	99.17	19.25	99.25	19.30	99.30
3	10.13	34.12	9.55	30.82	9.28	29.46	9.12	28.71	9.01	28.24
4	7.71	21.20	6.94	18.00	6.59	16.69	6.39	15.98	6.26	15.52
5	6.61	16.26	5.79	13.27	5.41	12.06	5.19	11.39	5.05	10.97
6	5.99	13.75	5.14	10.92	4.76	9.78	4.53	9.15	4.39	8.75
7	5.59	12.25	4.74	9.55	4.35	8.45	4.12	7.85	3.97	7.46
8	5.32	11.26	4.46	8.65	4.07	7.59	3.84	7.01	3.69	6.63
9	5.12	10.56	4.26	8.02	3.86	6.99	3.63	6.42	3.48	6.06
10	4.96	10.04	4.10	7.56	3.71	6.55	3.48	5.99	3.33	5.64
11	4.84	9.65	3.98	7.21	3.59	6.22	3.36	5.67	3.20	5.32
12	4.75	9.33	3.89	6.93	3.49	5.95	3.26	5.41	3.11	5.06
13	4.67	9.07	3.81	6.70	3.41	5.74	3.18	5.21	3.03	4.86
14	4.60	8.86	3.74	6.51	3.34	5.56	3.11	5.04	2.96	4.69
15	4.54	8.68	3.68	6.36	3.29	5.42	3.06	4.89	2.90	4.56
16	4.49	8.53	3.63	6.23	3.24	5.29	3.01	4.77	2.85	4.44
17	4.45	8.40	3.59	6.11	3.20	5.18	2.96	4.67	2.81	4.34
18	4.41	8.29	3.55	6.01	3.16	5.09	2.93	4.58	2.77	4.25
19	4.38	8.18	3.52	5.93	3.13	5.01	2.90	4.50	2.74	4.17
20	4.35	8.10	3.49	5.85	3.10	4.94	2.87	4.43	2.71	4.10
21	4.32	8.02	3.47	5.78	3.07	4.87	2.84	4.37	2.68	4.04
22	4.30	7.95	3.44	5.72	3.05	4.82	2.82	4.31	2.66	3.99
23	4.28	7.88	3.42	5.66	3.03	4.76	2.80	4.26	2.64	3.94
24	4.26	7.82	3.40	5.61	3.01	4.72	2.78	4.22	2.62	3.90
25	4.24	7.77	3.39	5.57	2.99	4.68	2.76	4.18	2.60	3.85
26	4.23	7.72	3.37	5.53	2.98	4.64	2.74	4.14	2.59	3.82
27	4.21	7.68	3.35	5.49	2.96	4.60	2.73	4.11	2.57	3.78
28	4.20	7.64	3.34	5.45	2.95	4.57	2.71	4.07	2.56	3.75
29	4.18	7.60	3.33	5.42	2.93	4.54	2.70	4.04	2.55	3.73
30	4.17	7.56	3.32	5.39	2.92	4.51	2.69	4.02	2.53	3.70
40	4.08	7.31	3.23	5.18	2.84	4.31	2.61	3.83	2.45	3.51
80	3.96	6.96	3.11	4.88	2.72	4.04	2.49	3.56	2.33	3.26
120	3.92	6.85	3.07	4.79	2.68	3.95	2.45	3.48	2.29	3.17
∞	3.84	6.63	3.00	4.61	2.60	3.78	2.37	3.32	2.21	3.02

	$n=6$		$n=8$		$n=12$		$n=24$		$n=\infty$	
m	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
1	233.99	5858.99	238.88	5981.07	243.91	6106.32	249.05	6234.63	254.31	6365.86
2	19.33	99.33	19.37	99.37	19.41	99.42	19.45	99.46	19.50	99.50
3	8.94	27.91	8.85	27.49	8.74	27.05	8.64	26.60	8.53	26.13
4	6.16	15.21	6.04	14.80	5.91	14.37	5.77	13.93	5.63	13.46
5	4.95	10.67	4.82	10.29	4.68	9.89	4.53	9.47	4.36	9.02
6	4.28	8.47	4.15	8.10	4.00	7.72	3.84	7.31	3.67	6.88
7	3.87	7.19	3.73	6.84	3.57	6.47	3.41	6.07	3.23	5.65
8	3.58	6.37	3.44	6.03	3.28	5.67	3.12	5.28	2.93	4.86
9	3.37	5.80	3.23	5.47	3.07	5.11	2.90	4.73	2.71	4.31
10	3.22	5.39	3.07	5.06	2.91	4.71	2.74	4.33	2.54	3.91
11	3.09	5.07	2.95	4.74	2.79	4.40	2.61	4.02	2.40	3.60
12	3.00	4.82	2.85	4.50	2.69	4.16	2.51	3.78	2.30	3.36
13	2.92	4.62	2.77	4.30	2.60	3.96	2.42	3.59	2.21	3.17
14	2.85	4.46	2.70	4.14	2.53	3.80	2.35	3.43	2.13	3.00
15	2.79	4.32	2.64	4.00	2.48	3.67	2.29	3.29	2.07	2.87
16	2.74	4.20	2.59	3.89	2.42	3.55	2.24	3.18	2.01	2.75
17	2.70	4.10	2.55	3.79	2.38	3.46	2.19	3.08	1.96	2.65
18	2.66	4.01	2.51	3.71	2.34	3.37	2.15	3.00	1.92	2.57
19	2.63	3.94	2.48	3.63	2.31	3.30	2.11	2.92	1.88	2.49
20	2.60	3.87	2.45	3.56	2.28	3.23	2.08	2.86	1.84	2.42
21	2.57	3.81	2.42	3.51	2.25	3.17	2.05	2.80	1.81	2.36
22	2.55	3.76	2.40	3.45	2.23	3.12	2.03	2.75	1.78	2.31
23	2.53	3.71	2.37	3.41	2.20	3.07	2.01	2.70	1.76	2.26
24	2.51	3.67	2.36	3.36	2.18	3.03	1.98	2.66	1.73	2.21
25	2.49	3.63	2.34	3.32	2.16	2.99	1.96	2.62	1.71	2.17
26	2.47	3.59	2.32	3.29	2.15	2.96	1.95	2.58	1.69	2.13
27	2.46	3.56	2.31	3.26	2.13	2.93	1.93	2.55	1.67	2.10
28	2.45	3.53	2.29	3.23	2.12	2.90	1.91	2.52	1.65	2.06
29	2.43	3.50	2.28	3.20	2.10	2.87	1.90	2.49	1.64	2.03
30	2.42	3.47	2.27	3.17	2.09	2.84	1.89	2.47	1.62	2.01
40	2.34	3.29	2.18	2.99	2.00	2.66	1.79	2.29	1.51	1.80
80	2.21	3.04	2.06	2.74	1.88	2.42	1.65	2.03	1.32	1.49
120	2.18	2.96	2.02	2.66	1.83	2.34	1.61	1.95	1.25	1.38
∞	2.10	2.80	1.94	2.51	1.75	2.18	1.52	1.79	1.00	1.00

Références

1. J.-R. Barra. *Notions fondamentales de statistiques mathématiques*. Dunod, 1971.
2. P. Bickel and K. Doksum. *Mathematical statistics*. Prentice Hall, 1977.
3. A. Borovkov. *Mathematical statistics*. Gordon & Breach Science Pub., 1999.
4. N. Bouleau. *Probabilités de l'ingénieur. Variables aléatoires et simulation*. Actualités Scientifiques et Industrielles, 1986.
5. P. Brémaud. *Introduction aux probabilités*. Springer Verlag, 1984.
6. D. Dacunha-Castelle and M. Duflo. *Probabilités et statistiques : Problèmes à temps fixe*. Masson, 1982.
7. W. Feller. *An introduction to probability theory and its applications*, volume 1. Wiley and Sons, Inc., 3rd edition, 1968.
8. G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford University Press, 2nd edition, 1992.
9. B. Jourdain. *Probabilités et statistique*. Ellipses, 2009.
10. A. Monfort. *Cours de statistique mathématique*. Economica, 1982.
11. A. Monfort. *Introduction à la statistique*. École Polytechnique, 1988.
12. J. Neveu. *Bases mathématiques du calcul des probabilités*. Masson, 1964.
13. J. Neveu. *Introduction aux probabilités*. École Polytechnique, 1988.
14. W. Rudin. *Real and complex analysis*. Mc Graw-Hill, 3rd edition, 1986.
15. G. Saporta. *Probabilités, analyse des données et statistique*. Technip, 2ème edition, 2006.
16. L. Schwartz. *Analyse III : calcul intégral*. Hermann, 1993.

Index

σ-additivité, 5

anniversaire, 10
arrangement, 10

Bayes (formule de), 12, 96
Bernoulli (schéma de), 32
Berry-Esséen (inégalité de), 152
Bertrand (paradoxe de), 21
biais, 206
Bonferroni (inégalités de), 19
Box-Muller (transformation de), 110

Cauchy-Schwarz (inégalité de), 42
changement de variable, 101
Cochran (théorème de), 174, 250
collectionneur, 71
convergence
  L^2, 130
  absolument, 62
  dominée, 62, 100, 127
  loi, 132
  monotone, 6, 101, 128
  presque sûre, 126
  probabilité, 51, 128
  rayon, 63
covariance, 48
  asymptotique, 222
  matrice, 167, 179

dénombrable, 15
dénombrement, 10
densité, 80
déviations
  grandes, 154
  modérées, 154
discernabilité, 9

écart type, 47
échantillon, 195
Edgeworth (développement d'), 155
entropie, 76, 163, 202
équidistribué, 138
erreur
  de 1$^{\text{ère}}$ espèce, 237
  de 2$^{\text{ème}}$ espèce, 237
espace
  mesurable, 16
  probabilisé, 5
  produit, 17
espérance, 40, 84
  conditionnelle, 57, 96
estimateur, 197
  amélioré, 216
  asymptotiquement normal, 222
  asymptotiquement efficace, 227
  asymptotiquement préférable, 226
  consistant, 197
  convergent, 197
  efficace, 211, 220
  faiblement convergent, 197
  fortement convergent, 198
  inadmissible, 205
  maximum de vraisemblance, 200, 201
  moindres carrés, 249, 253
  moments, 198, 223
  optimal, 217
  préférable, 205
  régulier, 210
  sans biais, 206

substitution, 198, 223
évènement, 4

Fatou (lemme de), 101, 128
FDCR, 210
fonction
caractéristique, 114
génératrice, 52, 114
intégrable, 99
pivotale, 294, 303
répartition, 81, 82, 137
empirique, 141, 276
risque quadratique, 205
formule
binôme, 10
crible, 6, 19
décomposition, 6, 12, 29, 46
Fréchet-Darmois-Cramer-Rao (borne de), 210
fréquence empirique, 141, 274
fractile, 83
Fubini (théorème de), 62, 99

Glivenko-Cantelli (théorème de), 142

histogramme, 144
hypothèse
alternative, 236
bilatérale, 246
composite, 244
explicite, 264
implicite, 260, 261
nulle, 236
simple, 239
unilatérale, 244

i.i.d., 138
indépendance
évènements, 13
variables aléatoires, 31, 48, 116
indicatrice, 28, 100
information de Fisher, 208
intervalle de confiance, 147, 151, 178, 294

jacobien, 102
Jensen (inégalité de), 42, 68

Kolmogorov-Smirnov (théorème de), 277

Lévy (théorème de), 133
Lebesgue (mesure de), 5, 16, 32, 82, 99
Lehman-Sheffé (théorème de), 217

LFGN, 138
log-vraisemblance, 200
loi, 26, 27
évènements rares, 37
Bernoulli, 28, 118
béta, 90, 198
binomiale, 34, 118
binomiale négative, 72
Cauchy, 88, 118
conditionnelle, 56, 96
Dirichlet, 215
exponentielle, 87, 118
faible des grands nombres, 51
Fisher-Snedecor, 178, 251
forte des grands nombres, 138
gamma, 89, 118
gaussienne, 85, 118, 212, 221
géométrique, 35, 118
Gumbel, 158
hypergéométrique, 160
χ^2, 90, 173, 267
marginale, 29, 31, 83
normale, 85
Poisson, 22, 37, 118
produit, 31, 196
sans mémoire, 69, 87, 108
Student, 176
uniforme, 28, 85, 118, 201, 225

Markov (inégalité de), 67
Méré (Chevalier de), 67
mesurable
ensemble, 16
espace, 16
fonction, 17
méthode du rejet, 99, 111
modèle
échantillonnage, 195
exponentiel, 218, 244
identifiable, 196
paramétrique, 196
régulier, 209
Monte-Carlo (méthode de), 139
de Montmort (problème de), 22
moyenne empirique, 51, 138

négligeable, 5
Neyman (principe de), 238
Neyman et Pearson, 239
niveau

asymptotique, 147, 151, 255, 298
d'un test, 238
exact, 178, 293
excès, 296, 304
noyau, 142

p-valeur, 242
asymptotique, 256
partition, 15
permutation, 10
presque partout, 100
presque sûr, 5
p.p., 100
p.s., 5
probabilité, 4
composée, 19
conditionnelle, 12
produit, 14
uniforme, 8

quantile, 83

Rao-Blackwell (amélioré de), 216
rapport de vraisemblance, 239
région
critique, 236, 301
de confiance, 293
régresseur, 248
régression linéaire, 248
résidus, 249
risque quadratique, 205

score, 207
sensibilité, 20
simulation, 97
Slutsky (théorème de), 149
sondage, 162
spécificité, 20
stabilisation de la variance, 181
statistique, 197
canonique, 218, 244
de test, 242
exhaustive, 214
minimale, 216, 219
totale, 215
Stirling, 23, 90, 221

table d'analyse de la variance, 252
Tchebychev (inégalité de), 42
TCL, 145
test, 260
adéquation, 267, 270, 275, 278, 279
aléatoire, 237
asymptotique, 255
convergent, 255
Hausman, 266
indépendance, 271
χ^2, 267, 270, 271
Kolmogorov-Smirnov, 275, 278
Neyman, 239
niveau, 238
niveau asymptotique, 255
puissance, 238
pur, 236
sans biais, 246
seuil, 238
UPP, 238
UPPS, 246
utilité des régresseurs, 251
Wald, 262, 281
théorème central limite, 145, 179
théorème de factorisation, 215
tirage
avec remise, 70
sans remise, 70, 71, 160
tribu, 16

variable aléatoire
continue, 82
définition, 26
densité, 82
discrète, 27
finie, 26
indépendante, 31, 48, 116, 169
intégrable, 40, 85
variance, 47
asymptotique, 222
empirique, 151, 175
vraisemblance, 200

Wald (équation de), 60
Weierstrass (théorème de), 75
Weldon (dés de), 290

zone de rejet, 236

Ouvrages déjà parus dans la même collection

Queiros-Conde, Diogo, Feidt, Michel, *Constructal theory and multi-scale geometries : Theory and applications in energetics, chemical engineering and materials*, Les Presses de l'ENSTA (Les actes), Paris, 2010, ISBN 978-2-7225-0921-4.

Ameisen, Jean-Claude, Klein, Etienne, Leglu, Dominique, *Écrire la science*, Les Presses de l'ENSTA (Les actes), Paris, 2010, ISBN 978-2-7225-0924-5.

Fieux, Michel, *L'océan planétaire*, Les Presses de l'ENSTA (Les cours), Paris, 2010, ISBN 978-2-7225-0915-3.

Rimond, Patrick, *Portraits*, Les Presses de l'ENSTA (Les invitations), Paris, 2010, ISBN 978-2-7225-0920-7.

Hilaire, Bruno, *Trente-deux*, Les Presses de l'ENSTA (Les invitations), Paris, 2010, ISBN 978-2-7225-0919-1.

Rosenbaum, Alexis, *Leçons d'introduction à la philosophie des sciences*, Les Presses de l'ENSTA (Les cours), Paris, 2009, ISBN 978-2-7225-0918-4.

Bovis, Alain, *Hydrodynamique navale : théories et modèles*, Les Presses de l'ENSTA (Les cours), Paris, 2009, ISBN 978-2-7225-0916-0.

Ciarlet, Patrick, Lunéville, Eric. *La méthode des éléments finis. Tome 1, Concepts généraux*, Les Presses de l'ENSTA (Les cours), Paris, 2009, ISBN 978-2-7225-0917-7.

Perez, Jérôme, *Théorie des champs classiques*, Les Presses de l'ENSTA (Les cours), Paris, 2008, ISBN 978-2-7225-0913-9.

Perez, Jérôme, *Gravitation classique : problème à N corps, de 2 à l'infini*, Les Presses de l'ENSTA (Les cours), Paris, 2008, ISBN 978-2-7225-0914-6.

Ouvrages à paraître dans la même collection

Ciarlet Patrick, Lunéville, Eric [et al.], *La méthode des éléments finis, Tome 2*, Les Presses de l'ENSTA (Les cours), Paris, 2010, ISBN 978-2-7225-0923-8.

Bovis, Alain, *Hydrodynamique navale : le sous-marin*, Les Presses de l'ENSTA (Les cours), Paris, 2010, ISBN 978-2-7225-0925-2.

ENSTA

Couverture : ENSTA/Service édition